

STIC-Biotech/ChemLib

88508

From: Yu, Misook
Sent: Saturday, March 08, 2003 1:08 PM
To: STIC-Biotech/ChemLib
Subject: 09/502,945

Please search proteins encoded by SEQ ID NOs 1-5. Thank you,

Examiner Misook Yu, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)

RECEIVED
MAR 10 2003
STIC

Edward Han
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 3/11/03
Date Completed: 3/25/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH: 5 - reverse to FA
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 03
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run On: March 21, 2003, 12:20:42 ; Search time 98.9379 Seconds
(without alignments)
12016.549 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346

Sequence: 1 ggaattcctctgtcgaagt.....aaacaaaagttaaaattt 2885

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p.model -DEV=xl
-Q=/cgn2_1/USPTO.spool/US09502945/runat_14032003_101058_19113/app_query.fasta_1.10979
-DB=SPREMBL_21 -OFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09502945@cgn_1_1565_@runat_14032003_101058_19113 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1831.5	34.3	878	4 Q9UFU7	Q9ufu7 homo sapien

ID	Q9UFU7	PRELIMINARY:	PRT:	878 AA.
AC	Q9UFU7			
DT	01-MAY-2000 (TRENBLrel. 13, Created)			
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TRENBLrel. 19, Last annotation update)			
DE	Hypothetical 94.3 kDa protein (Fragment).			
GN	DKP2P586J0917			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=UTERUS;			
RA	Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;			
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL117455; CAB55935.1;			
DR	InterPro: IPR000286; His_deacetylase.			
DR	Pfam: PF00850; Hist_deacetyl; 1.			
DR	PRINTS; PR01270; HDASUPER.			
KW	Hypothetical protein.			
FT	NON_TER 1			
SQ	SEQUENCE 878 AA; 94330 MW; 26882E693960E10D CRC64;			

ALIGNMENTS

RESULT 1

Q9UFU7
ID Q9UFU7
AC Q9UFU7
DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)

DE Hypothetical 94.3 kDa protein (Fragment).

GN DKP2P586J0917

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=UTERUS;

RA Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL117455; CAB55935.1;

DR InterPro: IPR000286; His_deacetylase.

DR Pfam: PF00850; Hist_deacetyl; 1.

DR PRINTS; PR01270; HDASUPER.

KW Hypothetical protein.

FT NON_TER 1

SQ SEQUENCE 878 AA; 94330 MW; 26882E693960E10D CRC64;

Alignment Scores:

Pred. No.: 2,55e-102 Length: 878
Score: 1831.50 Matches: 424
Percent Similarity: 56.03% Conservativeness: 124
Best Local Similarity: 43.35% Mismatches: 249
Query Match: 34.26% Indels: 182
DB: 4 Gaps: 22

US-09-502-945-2 (1-2885) x Q9UFU7 (1-878)

```
Qy 11 TTGTGGAAGTCAAGAGAGCCACAGCCGCGCTCAACCAATTCCTCCACAGACACCC 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27 LeuGluArgThrValHisProAsnSerProGlyIleProTyrArgThrLeuGluProLeu 46
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 71 AAATGCTGGGAGCCACCATCTCTTTGGACACAGAGTTCCTCCCGCCAGAGGGCCCC 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 47 GluThrGluGlyAlaThrArgSerMetLeuSerSerPheLeuProProValProSerLeu 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 131 CTGGGAGCGCTCCCTCTCAAACTGCTTTGCTGGGCGCTACGACAGTCGAGACGAC 190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 ProSerAspProProGluHis 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 191 TTCCTCCCTCCGAAACAGCCCTCTGAACCACTTGAAGTCCGTCAAGGCTAAACAG 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 74 PheProLeuArgGlyThrValSerGluProAsnLeuLysLeuArgTyrLysProLys 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 251 AAGTGGCTGAGCGGAGAGAGTCCCTCTCGCTGCGCAAGATGGGACTGTTATTAGC 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93 LysSerLeuGluArgGlyAsnProLeuLeuArgLysGluSerAlaPro 110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 311 ACCTTTAAGAGAGAGCTGTTGAGATCAAGGTGCGCGGCGCTGCGCTGCTGTGT 370
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 SerLeuArgArgProAlaGluThrLeuGlyAspSerSerProSerSerSer 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 371 ACAGGGACCGGCTCGGCGCCAGCTCTCCACAGAGTCCACAGACACCATCCTGAG 430
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 ---SerThrProAlaSerGlyCysSerProAsnAspSerGluHis 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 431 AATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATCCTCCCTCAGCACCGAGC 490
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 ---GlyProAsnProIleLeuGlySerGluAlaLeuLeuGlyGlnArgLeu 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 491 CTCCTCTGACAGCTCCGCCAACAGTTCAGCTCTACAGCTCTCTCTCTGCGCCAC 550
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 ArgLeuGlnGluThrSerValAlaProPheAlaLeuProThrValSerLeuProAla 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 551 ATCTCCTAGGCTGAGCGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 610
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 IleThrLeuGlyLeuProAla 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 611 AAGCTGTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 670
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 ---ProAlaArgAlaAspSerArgArgThrHisProThrLeuGlyProArgGly 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 671 AGCTGACCGGCAAGTTCATGACACATCTCTATTCTGCTGCTGCTGCTGCTGCTG 730
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 ProIleLeuGlySerProHisThrProLeuPheLeuProHis 221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 731 CTGAGGGGCGAGGAGCGCCCGCAGCGGATGCTTCCCTGCTGAGCATGTGCTGTG 790
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 222 ProGluAlaGlyGlyThr 238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 791 GAGCAGCGCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 850
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 AspProSerGlySerHisAlaProLeuLeuThrValProGlyLeuGlyProLeuProPhe 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 851 GTGACGGGTGAACGTGTGCGCCACAGCATGCGGAGCGGTAGGCAAGCTCCCGGCGATCG 910
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 259 HisPheAlaGlnSerLeuMetThrThrGluArgLeuSerGlySer 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 911 CCGCTGAGCGCGACTCAGTCCCTCAGCGCTGCGCAGAGTCCCGAGGCC 958
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 278 ProLeuSerArgThrArgSerGluProLeuProSerAlaThrAlaProProProPro 297
Qy 959 -----CTGCAGCAGCTGCTCATGCAACACAGCAGCAGCAGTTCCTG 1000
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 GlyProMetGlnProArgLeuGluGlnLeu -----LysThrHisValGlnValIle 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1001 GAGAAGCAGAGCAGCAGCAGCTACAGCTGGGCAAGATCTCACCAGAGCAGGGGAGCTG 1060
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 LysArgSer -----AlaLysProSerGluLys 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1061 CCCAGGAGCCACCACCCTGAGGAGCAGAGGAGGAGCTGAGGAGCAGCAGGAG 1120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 324 ProArg ----- 325
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1121 GTCCTGCTGGGAGGAGGAGCCCTGACCATGCCCGGAGGAGGCTCCACAGAGTGTAGAGC 1180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 ----- 325
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1181 ACAGAGAACACCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 326 LeuArgGlnIleProSerAlaGluAspLeuGluThrAspGly ----- 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1241 ATCCAGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 -----GlyGlyProGlyGlnValValAspAspGlyLeuGluHis 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1295 GAGGAGCTGCTGCTGATACAAAACTGTTCTAGATCCCAACCGCTGCAACCTTTG 1354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 ArgGluLeuGlyHisGlyGln -----ProGluAlaArgGlyProAlaProLeu 368
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1355 CAGGTGTACCAAGCGCCC ----- 1372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 Gln -----GlnHisProGlnValLeuLeuTrpGluGlnArgLeuAlaGlyArgLeu 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1373 -----CTCAGCTGGCCACTGTGCCCCCAGCAGCC 1402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 387 ProArgGlySerThrGlyAspThrValLeuLeuTrpGluGlnArgLeuAlaGlyArgLeu 406
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1403 CTGGGCGGTACCAATCTCCCTGCTGCGGCGGCGATGAAGAACCC ----- 1453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 407 LeuSerArgAlaGlnSerSerProAlaAlaProAlaSerLeuSerAlaProGluProAla 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1454 -----CCAGACCAACCCCTCAAGCACCTC ---TTCACCACA 1486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 SerGlnAlaArgValLeuSerSerSerGluThrProAlaArgThrLeuProPheThr 446
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1487 AGTGTGCTTACGACAGCTTCTAAAGCAGCAGTGTGCGGGAACACACACAGCTG 1546
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 447 GlyLeuIleThrAspSerValMetLeuLysHisGlnCysSerCysGlyAspAsnSerArg 466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1547 CACCTGAGCATGCTGGCGGATCCAGAGCATCTGGTCCCGGCTGAGGAGACAGGCTG 1606
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 467 HisProGluHisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnGluArgGlyLeu 486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1607 CTTAGCAAGTCCGAGCGGATCCGAGGTCGCAAGCCAGCTAGATGATGATCCAGACAGTG 1666
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 ArgSerGlnCysGluCysLeuArgGlyArgLysAlaSerLeuGluGlnLeuSerVal 506
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1667 CACTCTGAATACCACACCTCTCTATGGGACAGTCCCTCAACCGCAGAGAGCTAGAC 1726
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 HisSerGluArgHisValLeuLeuIleThrAsnProLeuSerArgLeuLysLeuAsp 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1727 AGCAAGAGTGTCTCGGTCCCATCAGCCAGAGATGTATGTCTGCTGCTGCTGCTG 1786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 527 AsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPheValMetLeuProCysGlyGly 546
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1787 ATCGGGGTGGAGTGACACCGTGTGGATGAGATGAGTCTCTCCAGTGTGCTCGGATG 1846
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 ValGlyValAspThrAspThrIleTrpAsnGluLeuHisSerSerAsnAlaAlaArgTrp 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1847 GCAGTGGCTGCTGCTGAGCTGCTTCAAGTGTGCTGAGGAGGAGCTCAACAATGGA 1906
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 567 AlaAlaGlySerValThrAspLeuAlaPheLysValAlaSerArgGluLeuLysAsnGly 586
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

QY 1907 TTTGCCATCATCGCGCCCGCCAGGACACACCGCGAGGAATCCACAGCCATGGGATCTGC 1966
Db 1908 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
587 PheAlaValAlaArgProProGlyHisHisAlaAspHisSerThrAlaMetGlyPheCys 606
QY 1967 TTTCTCAACTCTGTAGCAGCATCACCGCAAACTCTACAGCAGAGTGAACGTGGGCAAG 2026
Db 1968 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
607 PhePheAsnSerValAlaIleAlaCysArgGlnGlnGlnGlnGlnSerLysAlaSerLys 626
QY 2027 GTCCTCATCTGGACTGGGACATTCACATGGCAATGGCAATGGCAGCAGGCGTCTACAAT 2086
Db 2028 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
627 IleLeuIleValAspTAspAspValHisHisGlyAsnGlyThrGlnGlnThrPheTyrGln 646
QY 2087 GACCCCTCTGCTACATCTCTCCATCGCTATGACACGGGAACTTCTTTCAGGC 2146
Db 2088 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
647 AspProSerValLeuTyrIleSerLeuHisArgHisAspAspGlyAsnPhePheProGly 666
QY 2147 TCTGGGCTCTCTGAAGAGTTGGTGGAGGACGAGGCGTGGGTACATGTGAACGTGGCA 2206
Db 2148 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
667 SerGlyAlaValAspGluValGlyAlaGlySerGlyGlyGlyPheAsnValAlaValAla 686
QY 2207 TGGACAGGAGTGGACCCCGCCATGGAGAGCTGGAGTACCTTACAGCCTTCAGGACA 2266
Db 2208 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
687 TrpAlaGlyGlyLeuAspProProMetGlyAspProGluTyrLeuAlaAlaPheArgIle 706
QY 2267 GTGCTGATGCCATTGCCAGAGTTCTACCTGATGTGTCTAGTCTCCGCGGGTTT 2326
Db 2268 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 ValValMetProIleAlaArgGluPheSerProAspLeuValLeuValSerAlaGlyPhe 726
QY 2327 GATGCTCTTGAAGACATCTGTCTCTGGTGGTCTCTACCTGAGGCGGCGGTGGTGG 2386
Db 2328 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 AspAlaAlaGlyGlyHisProAlaProLeuGlyGlyTyrHisValSerAlaLysCysPhe 746
QY 2387 GGCATTGACAGGAGCTGATGACCTGGCAGGCGGCGGTGGTGGCGCTGGAG 2446
Db 2388 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
747 GlyTyrMetThrGlnGlnLeuMetAsnLeuAlaGlyGlyAlaValValLeuAlaLeuGlu 766
QY 2447 GGAGGCGATGACCTGACCGCATCTGTGATGCTCTGAGCTGTGCTCGGCTCTGCTC 2506
Db 2448 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
767 GlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValAlaAlaLeuLeu 786
QY 2507 AGTGTAAAGCTGAGCCCTTGGATGAGCAGTCTTGACGAAAGCCCAACATCAACGCA 2566
Db 2508 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
787 GlyAsnArgValAspProLeuSerGluGluGlyTrpLysGlnLysProAsnLeuAsnAla 806
QY 2567 GTGGCCACGCTAGAGAAATCATCGATCCAGATCCAGCAACACTGGAGCTGTGTCAGAA 2626
Db 2568 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
807 IleArgSerLeuGluAlaValIleArgValHisSerLysTyrTrpGlyCysMetGlnArg 826
QY 2627 TTCGCGCTGCTGCGCGGTCCCTGCGAGGCGGCCACAGGCTGAGACCGAGAGCC 2686
Db 2628 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
827 LeuAlaSerCysProAspSerTrpValProArgValProGlyAlaAspLysGluGluVal 846
QY 2687 GAAAT-GTGAACGCCATGGCTGTGCTGTGTTGGG-----GCCGAACAGGCCCCAAGCT 2739
Db 2688 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
847 GluAlaValThrAlaLeuAlaSerLeuSerValGlyIleLeuAlaGluAsp----- 863
QY 2740 CGGGCAGCCGGGAACACAGCCCCAGCGCGCAGAGAGCCCATGGAGCAGGAG 2793
Db 2741 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
864 -----ArgProSerGluGlnLeuValGluGluGlu 873

```

RESULT 2
Q9JL72

ID Q9JL72 PRELIMINARY; PRT; 938 AA.
AC
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Histone deacetylase 7.
GN HDAC7A OR HDAC7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6.C57BL/6;
RX MEDLINE=20107033; PubMed=10640276;
RA Kao H.Y., Downes M., Ordentlich P., Evans R.M.;
RT "Isolation of a novel histone deacetylase reveals that class I and
RL Genes Dev. 14:55-66(2000).
DR EMBL: AF207749; AAF31419.1;
DR MGD; MGI:1891835; Hdac7a.
DR InterPro; IPR000286; Hist_deacetylase.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PRO1270; HDASUPER.
SQ SEQUENCE 938 AA; 101318 MW; A282A36BEC840897 CRC64;

Alignment Scores:
Pred. No.: 8,69e-99 Length: 938
Score: 1773.00 Matches: 422
Percent Similarity: 54.79% Conservative: 110
Best Local Similarity: 43.46% Mismatches: 250
Query Match: 33.16% Indels: 190
DB: 11 Gaps: 24

US-09-502-945-2 (1-2885) x Q9JL72 (1-938)
QY 2 GAATTCCTCTTGTGGAAGTCAAGGAGGACACACAGGCGGCTCAACCATTCCTCCCA 61
Db 114 GluValIleLeuLysLysGlnGlnAlaLeuGluArgThrValHisProSerSerPro 133
QY 62 CAGCACCC-----AAATCTGGGGAGCCCAACCATCTCTTTTTC 100
Db 134 SerIleProTyrArgThrLeuGluProLeuAspThrGluGlyAlaAlaArgSerValLeu 153
QY 101 GACCAGAGTTCCTCCCGCCAGAGCGCCCTCGGAGCCCTCCCTCTACAACTGCCT 160
Db 154 SerSerPheLeuProValProSerLeuProThrGluProGluHis----- 170
QY 161 TTGCTGCGCCCTACGACAGTCCGAGCAGTTCCTCCCTCCGCAAAACAGCCTCTGAACCC 220
Db 171 -----PheProLeuArgLysThrValSerGluPro 180
QY 221 AACTTGAAGTCCGTTCAGGCTAAACAGAGGTGGCTGACGAGGAGAGAGTCCCTTC 280
Db 181 AsnLeuLysLeuArgTyrLysProLys---LysSerLeuGluArgLysAsnProLeu 199
QY 281 CTGCGTGCAGAGGATGGGACTCTATTAGCAGCTTTAAGAGAGAGCTGTTGAGATCACA 340
Db 200 LeuArgLysGluSerAlaPro-----ProSerLeuArgArgProAlaGluThrLeu 217
QY 341 GGTGCGGCGCTGGGCGCTGCTGCTGTGTAAACAGCGACCGCGCTCGGCGCCAGCTCT 400
Db 218 GlyAspSerSerProSerSerSer-----SerThrProAlaSerGlyCysSerSer 234
QY 401 CCCACAGCTCCACAGCAGCACCATCCTCGAGATGCTTACTTGCTAGTCCCAACATC 460
Db 235 ProAsnAspSer-----GluHisGly----- 241
QY 461 CCCACTGAGATCTCCTCCAGCAGCGAGCCCTCCCTCTGGACAGTCCCCCAACAGTTC 520
Db 241 ----- 241
QY 521 AGCCTCTACAGCTCTCTTCTGTCGCCAACATCTCCCTAGGCTGCAGGCCAGCGTCACT 580
Db 242 -----ProAsnProAlaLeuGlySerGluAlaAspGlyAsp 253
QY 581 GTCACCAACTCAGCAGCTCAGTCCCTCCCGAAG----- 613
Db 254 ---ArgArgThrHisSerThrLeuGlyProArgGlyProValLeuGlyAsnProHisAla 272
QY 614 ---CTGTGCGACACAGCAGGAGCGCGAGAGGAGCCCTCCAGTCCCTCGGCGCAGGTTGC 670
Db 273 ProLeuPheLeuHisHisGlyLeuGluProGluAla-----GlyGly 286

```


Db	925	ProSerGluArgLeuValGluGluGlu	933	:	::: :::
RESULT 3					
Q9NFK9		PRELIMINARY;	PRT;	855	AA.
ID	Q9NFK9				
AC	Q9NFK9				
DT	01-OCT-2000	(TrEMBLrel. 15, Created)			
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)			
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)			
DE	Histone deacetylase 7.				
GN	HDAC7.				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=CERVICAL CARCINOMA;				
RA	Li S., Fischle W., Verdin E., Walsh M.J.;				
RT	"A novel class II HDAC is associated with the transcriptional				
RT	homeodomain repressor CCAAT displacement protein.";				
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF239243; AAF63491.1;				
DR	InterPro; IPR000286; His_deacetylase.				
DR	Pfam; PF00850; Hist_deacetyl; 1.				
DR	PRINTS; PR01270; HDASUPER.				
SQ	SEQUENCE 855 AA; 92363 MW; 3157426AE1913DB6 CRC64;				
Alignment Scores:					
Pred. No.:	7.33e-93	Length:	855		
Score:	1674.50	Matches:	385		
Percent Similarity:	54.28%	Conservative:	103		
Best Local Similarity:	42.83%	Mismatches:	200		
Query Match:	31.32%	Indels:	211		
DB:	4	Gaps:	22		
us-09-502-945-2 (1-2885) x Q9NFK9 (1-855)					
QY	11	TTGTGCAAGTCAAGGAGCCACACACAGGGGGCTCAACATTCCTCCACAGCACCC	70		:
Db	101	LeuGluArgThrValHisProAsnSerProGlyIleProfyrArgThrLeuGluProLeu	120		:
QY	71	AAATGCTGGGAGCCACCATGCTCTTGGACCAGAGTTCCCTCCCCAGAGCGGCC	130		
Db	121	GluThrGluGlyAlaThrArgSerMetLeuSerSerPheLeuProValProSerLeu	140		
QY	131	CCTGGGAGCCTCCCTCCTACAAACTGCCCTTGCCTGGGGCCCTACGACAGTCGAGACG	190		
Db	141	ProSerAspProProGluHis	147		
QY	191	TTCCCCCTCGCAAAACAGCCTCTGAACCAACTGAAAGTGGTTCAAGGCTAAACAG	250		
Db	148	PheProLeuArgLysThrValSerGluProAsnLeuLysLeuArgfyrLysProLys	166		
QY	251	AAGTGGCTGAGCGGGAAGCAGTCCTCCTCGCTCGCAGGATCGGACTGTTATTAC	310		
Db	167	LysSerLeuGluArgArgLysAsnProLeuArgLysGluSerAlaPro	184		
QY	311	ACCTTTAAGAAGAGAGCTGTTGAGATCAAGGTGCGGGCCTGGGCGTCGTCGTGT	370		
Db	185	SerLeuArgArgArgProAlaGluThrLeuGlyAspSerSerProSerSerSer	202		
QY	371	ACAGCGCACCCGGCTCGGCGCCAGCTCTCCCAACAGCTCCCAACAGCACCATCGCTG	430		
Db	203	---SerThrProAlaSerGlyCysSerSerProAsnAspSer	216		
QY	431	AATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATGCTCCTCAGCACCGAGCC	490		
Db	217	HisGly	218		
QY	491	CTCCCTCTGGACAGTCCCCCAACCAAGTTTCAGCCTCTACACGCTCTCTCTGCGCCAA	550		


```
QY 1487 AGTGTGCTACGACACCTTCTATGCTAAAGCACCAGCTGTCATGTCGGGAACACACACGCTG 1546
Db 481 GlyLeuIleTyrAspSerValMetLeuLysHisGlnCysSerCysGlyAspAsnSerArg 500
QY 1547 CACCTGTAGCATGCTGCGGATCCAGACATCTGTCCCGCTGCGAGGAGACAGCCCTG 1606
Db 501 HisProGluHisAlaGlyArgIleGlnSerIleTyrSerArgLeuGlnGluArgGlyLeu 520
QY 1607 CTTAGCAAGTGGCGGATCCGAGTCCGAGCTCCAGACCTAGATGATGATCCAGACAGT 1666
Db 521 ArgSerGlnCysGluCysLeuArgGlyArgLysAlaSerLeuGluGlnLeuGlnSerVal 540
QY 1667 CACTCTGAATACACACCTGCTCTATGGACACAGTCCCTCAACCGCGGAGAGCTAGAC 1726
Db 541 HisSerGluArgHisValLeuLeuTyrGlyThrAsnProLeuSerArgLeuLysLeuAsp 560
QY 1727 AGCAAGAAGTGTGCTGCTCCATCAGCCAGAGATGTATGCTGTGCTGCTGCTGGGGC 1786
Db 561 AsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPheGluMetLeuProCysGlyGly 580
QY 1787 ATCGGGGTGGACAGTGCACACCGTGGTGAATGAGATGACCTCTCCAGTGTGTCGCATG 1846
Db 581 ValGlyValAspThrAspThrIleTyrPasnGlnLeuHisSerAsnAlaAlaArgTyr 600
QY 1847 GCAGTGGGCTGCTGCTGAGCTGGCTTCAAGTGTGCTGCTGAGAGAGCTCAAGAATGA 1906
Db 601 AlaAlaGlySerValThrAspLeuAlaPheLysValAlaSerArgGluLeuLysAsnGly 620
QY 1907 TTTCGCATCATCCGCGCCCGGACACCGCCGCGGAGATCCACAGCCATGGATCTGTC 1966
Db 621 PheAlaValValArgProProGlyHisHisAlaAspHisSerThrAlaMetGlyPheCys 640
QY 1967 TTCTTCAACTCTAGCATCACCGCAAACTCTACAGCAGAAG-----TTGAAC 2017
Db 641 PhePheAsnSerValAlaIleAlaCysArgGlnLeuGlnGlnSerLysAlaSerLys 660
QY 2018 GTGGGCAAGCTCTCATCTGGTGGACTGGGACATTCACCATGGCAATGCCACCCAGCGG 2077
Db 661 AlaSerLysIleLeuIleValAspTyrAspValHisHisGlyAsnGlyThrGlnGlnThr 680
QY 2078 TTCTACATGACCCCTCTGCTCTACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2137
Db 681 PheTyrGlnAspProSerValLeuTyrIleSerLeuHisArgHisAspAspGlyAsnPhe 700
QY 2138 TTTCAGGCTCTGGGCTCTCTGAAGAGTGTGGAGGACCGCGGCGGTGCAATGTG 2197
Db 701 PheProGlySerGlyAlaValAspGluValGlyAlaGlySerGlyGluGlyPheAsnVal 720
QY 2198 AACGTGGCATGGACAGGAGTGTGGACCCCGCCATGGAGACGTGGAGTACCTTACAGCC 2257
Db 721 AsnValAlaAlaTyrPalaGlyGlyLeuAspProProMetGlyAspProGluTyrLeuAlaAla 740
QY 2258 TTCAGGACAGTGTGATGCCCATTTGCCCGCATGTTCTCACCTGATGTTGCTAGTCTCC 2317
Db 741 PheArgIleValMetProIleAlaArgIlePheSerProAspLeuValLeuValSer 760
QY 2318 GCCGGTTTGTATGCTGTGAAGACATCTGTCTCTCTGGTGGCTACTCTGTCACCGCC 2377
Db 761 AlaGlyPheAspAlaAlaGluGlyHisProAlaProLeuGlyGlyTyrHisValSerAla 780
QY 2378 AGATGTTTTGGCCATTGTACACGACGATGATGACCTGGCGAGGGCGCGGTGGTCTG 2437
Db 781 LysCysPheGlyTyrMetThrGlnGlnLeuMetAsnLeuAlaGlyGlyAlaValValLeu 800
QY 2438 GCCCTGGAGGGAGCCATGATTCACCGCCATCTGTGATGCCCTGAGCTGTGCTCTG 2497
Db 801 AlaLeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValAla 820
QY 2498 GCTCTGCTCACTGTAAGCTGCACCCCTTGGATGAGGAGCTGTTCGACGAAAGCC 2554
Db 821 AlaLeuLeuGlyAsnArgValAspProLeuSerGluGluGlyTyrPlysGlnLysPro 839
```

```
RESULT 4
Q8WUI4
ID Q8WUI4 PRELIMINARY: PRT: 614 AA.
AC Q8WUI4: Q8B73:
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Histone deacetylase 7A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 339-614 FROM N.A.
RC TISSUE=B-CELL;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020505; AAH20505.1;
DR EMBL; BC06453; AAH06453.1;
DR InterPro; IPR000286; His_deacetylase.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PRO1270; HDASUPER.
SQ SEQUENCE 614 AA; 66186 MW; 2B638DD8C866B502 CRC64;

Alignment Scores:
Pred. No.: 7,41e-89 Length: 614
Score: 1607.50 Matches: 343
Percent Similarity: 61.32% Conservative: 85
Best Local Similarity: 49.14% Mismatches: 131
Query Match: 30.07% Indels: 140
DB: 4 Gaps: 14

US-09-502-945-2 (1-2885) x Q8WUI4 (1-614)
QY 851 GTGACGGTGAACGTGTGGCCACACGATGCGGAGCTAGGCAAGCTCCCGCGCATCGG 910
Db 1 MetThrThrGluArgLeuSerGlySer-----GlyLeuHisTyr 13
QY 911 CCCTGAGCGGCACTCAGTCTCCACGCTGCGCGCAGAGTCCCGCGCC----- 958
Db 14 ProLeuSerArgThrArgSerGluProLeuProSerAlaThrAlaProProPro 33
QY 959 -----CTGCAGCAGCTGGTTCATCAACAACAGCAGCAGTTCCTG 1000
Db 34 GlyProMetGlnProArgLeuGluGlnLeu-----LysThrHisValGlnValIle 50
QY 1001 GAGAAGCAAGACAGCAGCAGCTACAGCTGGCGCAAGATCTCACCAGAGAGAGGAGCTG 1060
Db 51 LysArgSer-----AlaLysProSerGluLys 59
QY 1061 CCAGGAGCCACACACCCCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120
Db 60 ProArg----- 61
QY 1121 GTCTTGTGGGGAGGAGCCCTGACCATGCCCGGAGGCTCCACAGAGAGAGAGAGAG 1180
Db 61 ----- 61
QY 1181 ACACAGAGACCTGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1240
Db 62 LeuArgGlnIleProSerAlaGluAspLeuGluThrAspGly----- 75
QY 1241 ATCCAGGTTAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1294
Db 76 -----GlyGlyProGlyGlnValValAspAspGlyLeuGluHis 88
QY 1295 GAGGAGCCTGTGCTGGATACAAAAACTTCTCTAGATGCCCAACCGCTGCAACCTTTG 1354
Db 1295 GAGGAGCCTGTGCTGGATACAAAAACTTCTCTAGATGCCCAACCGCTGCAACCTTTG 1354
```



```
Db 89 ArgGluLeuGlyHisGlyGln-----ProGluAlaArgGlyProAlaProLeu 104
QY 1355 CAGGTGTACCAAGCCGCC----- 1372
Db 105 Gln-----GlnHisProGlnValLeuLeuTrpGluGlnArgLeuAlaGlyArgLeu 122
QY 1373 -----CTCAGCCCTGGCCACTGTGTCGCCCAACCAAGCC 1402
Db 123 ProArgGlySerThrGlyAspThrValLeuLeuProLeuAlaGlnGlyHisArgPro 142
QY 1403 CTGGCCCTACCAATCTCCCTGCTGCCCTGGGGCATGAAGAACCCC----- 1453
Db 143 LeuSerArgAlaGlnSerProAlaAlaProAlaSerLeuSerAlaProGluProAla 162
QY 1454 -----CCACACCAACCCGTCAGCACCTC-----TTCCACCA 1486
Db 163 SerGlnAlaArgValLeuSerSerSerGluThrProAlaArgThrLeuProPheThrThr 182
QY 1487 AGTGTGGTCTACGACACGTTTCATGCTAAAGCACCATGTCATGTGCGGGAACACACACAGTG 1546
Db 183 GlyLeuIleTyrAspSerValMetLeuLysHisGlnCysSerCysGlyAspAsnSerArg 202
QY 1547 CACCTGAGCATGTGGCCGATCCAGAGCATCTGTGTCGCCGCTGCAGGACAGAGCGCTG 1606
Db 203 HisProGluHisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnArgGlyLeu 222
QY 1607 CTTAGCAAGTCGGAGCGATCCGAGTCCGAAGCCAGCTAGATGATCCAGACAGTG 1666
Db 223 ArgSerGlnCysGluCysLeuArgGlyArgLysAlaSerLeuGluGlnSerVal 242
QY 1667 CACTCTGAATACCACACCTCTCTATGGGACCATGTCCTCAACCGCAGCAAGCTAGAC 1726
Db 243 HisSerGluArgHisValLeuLeuTyrGlyThrAsnProLeuSerArgLeuLysLeuAsp 262
QY 1727 AGCAAGAGTTGCTGCTCCATCCAGCAGAGATGTATGCTGCTGCTGCTGCTGCTGCTGCTG 1786
Db 263 AsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPheValMetLeuProCysGlyGly 282
QY 1787 ATCCGGGTGGACATGACACCGTGTGAATGATGATGATGATGATGATGATGATGATGATG 1846
Db 283 ValGlyValAspThrAspThrIleTrpAsnGluLeuHisSerSerAsnAlaAlaArgTrp 302
QY 1847 GCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1906
Db 303 AlaAlaGlySerValThrAspLeuAlaPheLysValAlaSerArgGluLeuLysAsnGly 322
QY 1907 TTGCGCATCATCGGCCGCCAGGACACCCAGCCAGGAAATCCAGCATGGATTCGTC 1966
Db 323 PheAlaValValArgProProGlyHisHisAlaAspHisSerThrAlaMetGlyPheCys 342
QY 1967 TTCTTCACTCTGTAGCCATCACCAGCAAACTCTACAGCAGAAGTTGAACGTGGGCAAG 2026
Db 343 PhePheAsnSerValAlaIleAlaCysArgGlnLeuGlnSerLysAlaSerLys 362
QY 2027 GTCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2086
Db 363 IleLeuIleValAspTrpAspValHisHisGlyAsnGlyThrGlnGlnThrPheTyrGln 382
QY 2087 GACCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2146
Db 383 AspProSerValLeuTyrIleSerLeuHisHisHisAspGlyAsnPhePheProGly 402
QY 2147 TCTGGGCTCTCAAGAGTGTGGGAGCACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2206
Db 403 SerGlyAlaValAspGluValGlyAlaGlySerGlyGlyGlyPheAsnValAsnValAla 422
QY 2207 TGGACAGGAGTGTGGACCCCTTACCTGAGAGCTGGAGCTGAGTACCTTACAGCTTCAGACA 2266
Db 423 TrpAlaGlyLeuAspProMetGlyAspProGluTyrLeuAlaAlaPheArgIle 442
QY 2267 GTGGTATGCCCATTCGCCACAGTGTCTCACCTGATGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2326
Db 443 ValValMetProIleAlaArgGluPheSerProAspLeuValLeuValSerAlaGlyPhe 462
```

```
QY 2327 GATCCTGTTGAAGACATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2386
Db 463 AspAlaAlaGluGlyHisProAlaProLeuGlyGlyTyrHisValSerAlaLysCysPhe 482
QY 2387 GGCACATTCACAGGACGCTGATGACCTGGCAGGCGCGGGTGGTGGTGGTGGTGGTGGTGGTGG 2446
Db 483 GlyTyrMetThrGlnGlnLeuMetAsnLeuAlaGlyGlyAlaValValLeuAlaLeuGlu 502
QY 2447 GGAGCCATGACTGACCGCCATCTGTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2506
Db 503 GlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValAlaAlaLeuLeu 522
QY 2507 AGTGAAGCTGACCGCTTGGATGAGCAGTCTTGCAGCAAAAGCCCAACATCAACGCA 2566
Db 523 GlyAsnArgValAspProLeuSerGluGluGlyTrpLysGlnLysProAsnLeuAsnAla 542
QY 2567 GTGCCACGCTAGAGAAGTCATCGAGATCCAGAGCAACACCTGGAGCTGTGTGCAGAG 2626
Db 543 IleArgSerLeuGluAlaValIleArgValHisSerLysTyrTrpGlyCysMetGlnArg 562
QY 2627 TTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2686
Db 563 LeuAlaSerCysProAspSerTrpValProArgValProGlyAlaAspLysGluGluVal 582
QY 2687 GAAAT-GTGAACGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2739
Db 583 GluAlaValThrAlaLeuAlaSerLeuSerValGlyIleLeuAlaGluAsp----- 599
QY 2740 CGCGCAGCCCGGACACACAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2793
Db 600 -----ArgProSerGluGlnLeuValGluGluGlu 609

RESULT 5
Q9NW41 PRELIMINARY; PRT; 614 AA.
ID Q9NW41;
AC Q9NW41;
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA FLJ10328 fis, clone NT2RM2000588, weakly similar to histone
DE deacetylase HDAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001190; BAA91545.1;
DR InterPro; IPR000286; His_deacetylase.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PRO1270; HDASUPER.
SQ SEQUENCE 614 AA; 66173 MW; 36C82D314014F211 CRC64;
```

```
Alignment Scores:
Pred. No.: 7, 41e-89 Length: 614
Score: 1607.50 Matches: 343
Percent Similarity: 61.32% Conservative: 85
Best Local Similarity: 49.14% Mismatches: 131
Query Match: 30.07% Indels: 140
DB: 4 Gaps: 14
```

US-09-502-945-2 (1-2885) x Q9NW41 (1-614)

01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CG1770 protein.
 GN HDAC4 OR CG1770.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyroidae; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RP
 RN
 R1
 R2
 R3
 R4
 R5
 R6
 R7
 R8
 R9
 R10
 R11
 R12
 R13
 R14
 R15
 R16
 R17
 R18
 R19
 R20
 R21
 R22
 R23
 R24
 R25
 R26
 R27
 R28
 R29
 R30
 R31
 R32
 R33
 R34
 R35
 R36
 R37
 R38
 R39
 R40
 R41
 R42
 R43
 R44
 R45
 R46
 R47
 R48
 R49
 R50
 R51
 R52
 R53
 R54
 R55
 R56
 R57
 R58
 R59
 R60
 R61
 R62
 R63
 R64
 R65
 R66
 R67
 R68
 R69
 R70
 R71
 R72
 R73
 R74
 R75
 R76
 R77
 R78
 R79
 R80
 R81
 R82
 R83
 R84
 R85
 R86
 R87
 R88
 R89
 R90
 R91
 R92
 R93
 R94
 R95
 R96
 R97
 R98
 R99
 R100
 R101
 R102
 R103
 R104
 R105
 R106
 R107
 R108
 R109
 R110
 R111
 R112
 R113
 R114
 R115
 R116
 R117
 R118
 R119
 R120
 R121
 R122
 R123
 R124
 R125
 R126
 R127
 R128
 R129
 R130
 R131
 R132
 R133
 R134
 R135
 R136
 R137
 R138
 R139
 R140
 R141
 R142
 R143
 R144
 R145
 R146
 R147
 R148
 R149
 R150
 R151
 R152
 R153
 R154
 R155
 R156
 R157
 R158
 R159
 R160
 R161
 R162
 R163
 R164
 R165
 R166
 R167
 R168
 R169
 R170
 R171
 R172
 R173
 R174
 R175
 R176
 R177
 R178
 R179
 R180
 R181
 R182
 R183
 R184
 R185
 R186
 R187
 R188
 R189
 R190
 R191
 R192
 R193
 R194
 R195
 R196
 R197
 R198
 R199
 R200
 R201
 R202
 R203
 R204
 R205
 R206
 R207
 R208
 R209
 R210
 R211
 R212
 R213
 R214
 R215
 R216
 R217
 R218
 R219
 R220
 R221
 R222
 R223
 R224
 R225
 R226
 R227
 R228
 R229
 R230
 R231
 R232
 R233
 R234
 R235
 R236
 R237
 R238
 R239
 R240
 R241
 R242
 R243
 R244
 R245
 R246
 R247
 R248
 R249
 R250
 R251
 R252
 R253
 R254
 R255
 R256
 R257
 R258
 R259
 R260
 R261
 R262
 R263
 R264
 R265
 R266
 R267
 R268
 R269
 R270
 R271
 R272
 R273
 R274
 R275
 R276
 R277
 R278
 R279
 R280
 R281
 R282
 R283
 R284
 R285
 R286
 R287
 R288
 R289
 R290
 R291
 R292
 R293
 R294
 R295
 R296
 R297
 R298
 R299
 R300
 R301
 R302
 R303
 R304
 R305
 R306
 R307
 R308
 R309
 R310
 R311
 R312
 R313
 R314
 R315
 R316
 R317
 R318
 R319
 R320
 R321
 R322
 R323
 R324
 R325
 R326
 R327
 R328
 R329
 R330
 R331
 R332
 R333
 R334
 R335
 R336
 R337
 R338
 R339
 R340
 R341
 R342
 R343
 R344
 R345
 R346
 R347
 R348
 R349
 R350
 R351
 R352
 R353
 R354
 R355
 R356
 R357
 R358
 R359
 R360
 R361
 R362
 R363
 R364
 R365
 R366
 R367
 R368
 R369
 R370
 R371
 R372
 R373
 R374
 R375
 R376
 R377
 R378
 R379
 R380
 R381
 R382
 R383
 R384
 R385
 R386
 R387
 R388
 R389
 R390
 R391
 R392
 R393
 R394
 R395
 R396
 R397
 R398
 R399
 R400
 R401
 R402
 R403
 R404
 R405
 R406
 R407
 R408
 R409
 R410
 R411
 R412
 R413
 R414
 R415
 R416
 R417
 R418
 R419
 R420
 R421
 R422
 R423
 R424
 R425
 R426
 R427
 R428
 R429
 R430
 R431
 R432
 R433
 R434
 R435
 R436
 R437
 R438
 R439
 R440
 R441
 R442
 R443
 R444
 R445
 R446
 R447
 R448
 R449
 R450
 R451
 R452
 R453
 R454
 R455
 R456
 R457
 R458
 R459
 R460
 R461
 R462
 R463
 R464
 R465
 R466
 R467
 R468
 R469
 R470
 R471
 R472
 R473
 R474
 R475
 R476
 R477
 R478
 R479
 R480
 R481
 R482
 R483
 R484
 R485
 R486
 R487
 R488
 R489
 R490
 R491
 R492
 R493
 R494
 R495
 R496
 R497
 R498
 R499
 R500
 R501
 R502

Alignment Scores:	
Score No.:	2e-88
Score:	1601.50
Percent Similarity:	50.23%
Best Local Similarity:	36.60%
Query Match:	29.96%
DB:	5
Gaps:	35
Indels:	241
Mismatches:	304
Conservative:	149
Matches:	400
Length:	1235

US-09-502-945-2 (1-2885) x Q9VVF3 (1-1235)

QY 5 TTCTCTTTCG---AAGTCAAGGAGCCCCACACACCGGCGGCTCAACCATTCCTCCCA 61
 |||:::||||| |||||::: ||| ||| |||
 Db 171 PheIleuSerArgLysSerGlnAlaAlaAaSerAsnGlyThrThrThrThrSerPro 190

QY 823 TGTGCACTCCACGGGCAGTCCCCACTAGTGACGGGTGAACGTGTGGCCACCAGCAT-- 879

AC Q9H7L0;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE FLJ00062 protein (Fragment).
 GN FLJ00062.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SPLEEN;
 RA Ohara O., Nagase T., Kikuno R., Okumura K.;
 RT "The nucleotide sequence of a long cDNA clone isolated from human
 spleen.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK024469; BAB15759.1;
 DR InterPro; IPR000286; His_deacetylase.
 DR Pfam; PF00850; Hist_deacetyl; 1.
 DR PRINTS; PR01270; HDA\$UPER.
 FT NON_TER
 SO SEQUENCE 484 AA; 52327 MW; 1B1B0B363436A800 CRC64;

Alignment Scores:
 Pred. No.: 1-23e-83 Length: 484
 Score: 1520.50 Matches: 295
 Percent Similarity: 74.90% Conservative: 69
 Best Local Similarity: 60.70% Mismatches: 100
 Query Match: 28.44% Indels: 23
 DB: 4 Gaps: 5

US-09-502-945-2 (1-2885) x Q9H7L0 (1-484)

QY 1357 GGTGTACCAAGCGCCCTCAGCCTGGCCACTGTGCCCAACAGCCCTGGCGGTACCCA 1416
 Db 10 GlyValArgGlyValTyProProGlySerMetTrpValProAla-----ValAla 26

QY 1417 ATCTCTCCCTGCTGCT-----CCTGGGGGATGAAGAACCCTCCAGACCAACC 1464
 Db 27 ValLeuAlaCysSerLeuGlnProArgProTrp-GlyValArgThr---ProTrpValPr 45

QY 1465 CGTCAAGCACTCTTCAACCAAGTGGTCTAGGACACCTTCATGCTAAGACACAGTG 1524
 Db 45 oAlaLeuThrLeuAlaProAlaGlyLeuIleTyAspSerValMetLeuLysHisGlnCy 65

QY 1525 CATGTGGGGAACACACACGTGCACCTGAGCATGCTGGCGGATCCAGACGATCTGGTC 1584
 Db 65 sSerCysGlyAspAsnSerArgHisProGluHisAlaGlyArgIleGlnSerIleTrpSe 85

QY 1585 CCGGCTCAGGACAGCCCTGCTTAGCAAGTGGAGCGGATCCGAGGTCCGAAGCCAC 1644
 Db 85 rArgLeuGlnGluArgGlyLeuArgSerGlnCysGluCysLeuArgGlyArgLysAlaSe 105

QY 1645 GCTAGATGAGATCCAGACAGTGCACCTGAATACCAACACCTGCTGTATGGGACAGTCC 1704
 Db 105 rLeuGluGluLeuGlnSerValHisSerGluArgHisValLeuLeuTyroGlyThrAsnPr 125

QY 1705 CCTCAACGGCAGAGCTAGACAGCAAGAAGTTGCTGCTGCCATCGCCAGCAAGATGA 1764
 Db 125 oLeuSerArgLeuLysLeuAspAsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPh 145

QY 1765 TGCTGTGCTGCTTGGGGGATCGGGTGGACAGTGCACACCGTGTGGATGAGATGCA 1824
 Db 145 eValMetLeuProCysGlyGlyValGlyValAlaAspThrAspThrIleTrpAsnGluLeuHi 165

QY 1825 CTCTCTCAGTGTGCGGATGCGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1884
 Db 165 sSerSerAsnAlaAlaArgTrpAlaAlaGlyArgValThrAspLeuAlaPhelLysValAl 185

QY 1885 TGCAGGAGAGCTCAAGAATGATTGTCATATCCCGGCCCCAGGACACACCGCCAGGA 1944
 Db 185 aSerArgGluLeuLysAsnGlyPheAlaValValArgProProGlyHisHisAlaAspHi 205

QY 1945 ATCCACAGCCATGGGATTTCTTCAACTCTAGCCATCACCCGCAAACTCCTACA 2004
 Db 205 sSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIleAlaCysArgGlnLeuGl 225

QY 2005 GCAGAAAGTTGAACGTGGGCAAGTCTCATCTGCTGGAGTGGACATTCACCATGGCAATGG 2064
 Db 225 nGlnGlnSerLysAlaSerLysIleLeuIleValAspTrpAspValHisHisGlyAsnGl 245

QY 2065 CACCCAGCAGCGCTTCTACAAATGACCCCTCTGCTGCTCTACATCTCTCTGCTGCTATGA 2124
 Db 245 yThrGlnGlnThrPheTyrglnAspProSerValLeuTyIleSerLeuHisArgHisAs 265

QY 2125 CAACGGGAATCTTCTCCAGGCTCTGGGGTCTGAAGAGTGTGGAGGACGACCGGT 2184
 Db 265 pAspGlyAsnPhePheProGlySerGlyAlaValAspGluValGlyAlaGlySerGlyGl 285

QY 2185 GGGTCAATGTGAACGTGGCATGGACAGAGGTGGACCCCGCCCATTTGGAGACGTGA 2244
 Db 285 uGlyPheAsnValAsnValAlaTrpAlaGlyGlyLeuAspProProMetGlyAspProGl 305

QY 2245 GTACCTTACAGCCTTCAGGACAGTGTGATGCCATTGCCCACAGAGTCTCTACCTGATGT 2304
 Db 305 uTyrlLeuAlaAlaPheArgIleValValMetProIleAlaArgGluPheSerProaspLe 325

QY 2305 GTCTCTAGTCTCCGCGGGTTTGAATCTGTGAAGGACATCTGTCTCTCTGGTGGCTA 2364
 Db 325 uValLeuValSerAlaGlyPheAspAlaAlaGluGlyHisProAlaProLeuGlyGlyTy 345

QY 2365 CTCTGTACCGCCGATGTTTGGCCACTGTGACCGAGGAGTGTACCTGGCAGGGGG 2424
 Db 345 rHisValSerAlaLysCysPheGlyTyMetThrGlnGlnLeuMetAsnLeuAlaGlyGl 365

QY 2425 CCGGGTGTGCTGGCCCTGGAGGAGGACATGACTGTGACCGCCATCTGTGATGCCCTCTGA 2484
 Db 365 yAlaValValLeuAlaLeuGluGlyHisAspLeuThrAlaIleCysAspAlaSerGl 385

QY 2485 AGCTTGTCTCGGCTCTCTCAGTGTAAAGCTGACGAGCCCTTGGATGAGGACGATCTTGA 2544
 Db 385 uAlaCysValAlaAlaLeuLeuGlyAsnArgValAspProLeuSerGluGluGlyTrpLy 405

QY 2545 GCNAAGCCCAACATCAACGACGTCGACAGTGGCCACCGTAGAGAAGTCAAGAGATCCAGAGCAA 2604
 Db 405 sGlnLysProAsnLeuAsnAlaIleArgSerLeuGluAlaValIleArgValHisSerLy 425

QY 2605 ACACGTGAGCTGTGTCAGAAGTTGCGCGTGTGCTGGGCGGTCCTCGAGGGGCCCCA 2664
 Db 425 sTyrlTrpGlyCysMetGlnArgLeuAlaSerCysProAspSerTrpValProArgValPr 445

QY 2665 AGCAGTGTAGACCCGAAGAGCCGAAAT-GTGAACGCCCATGGCTTGTCTGTGGTGGGG-- 2721
 Db 445 oGlyAlaAspLysGluGluValGluAlaValThrAlaLeuAlaSerLeuSerValGlyTrl 465

QY 2722 ----GCCGAACAGGCCAGCTGCGGACCGCCGGGAACACACCCCGGCGGACAGGA 2777
 Db 465 eLeuAlaGluAsp-----ArgProSerGluGl 474

QY 2778 GCCCATGGAGCAGGAG 2793
 Db 474 nLeuValGluGluGlu 479

RESULT 8
 Q96DY4
 ID Q96DY4 PRELIMINARY; PRT; 363 AA.
 AC Q96DY4
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Unknown (Protein for IMAGE:3867625) (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;


```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EYE;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013140; AAH13140.1; -
FT NON_TER 1
SQ SEQUENCE 363 AA; 40197 MW; 1EF788B16A0C6187 CRC64;

Alignment Scores:
Pred. No.: 7,14e-69 Length: 363
Score: 1275.00 Matches: 245
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.85% Indels: 0
DB: 4 Gaps: 0

US-09-502-945-2 (1-2885) x Q96DY4 (1-363)
QY 2 GAATTCCTCTGTGCAAGTCAAGAGGAGCCACACACAGGCGGCTCAACCATTCCTCCCA 61
Db 116 GluPheLeuLeuSerLysLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 135
QY 62 CAGCACCCCAATGCTGGGAGCCACCATGCTTCTTTGGACCAGAGTTCCCTCCCAAG 121
Db 136 GluHisProLysCysTrpGlyAlaHisHisAlaSerLeuAspGlnSerSerProGln 155
QY 122 AGCGGCGCCCTGGGAGGCTCCTCTACAAACTGCCTTGGCTGGGCGCTACGACAGT 181
Db 156 SerGlyProGlyThrProProSerTyLysLeuProLeuProGlyProTyAspSer 175
QY 182 CGAGACGACTTCCCTCCCAAAACAGCCTCTGAACCAACTGAAAGTGCCTCAAGG 241
Db 176 ArgAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg 195
QY 242 CTAACACAGAGTGCTGAGCGGAGAGCAGTCCCTCTCGCTGCGAGGATGGGACT 301
Db 196 LeuLysGlnLysValAlaGluArgSerSerProLeuLeuArgLysAspGlyThr 215
QY 302 GTTATTAGCACCTTAAAGAGAGCTGTGAGATCACAGTCCGCGCTGGCGCTGCG 361
Db 216 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyAlaGlyProGlyAlaSer 235
QY 362 TCGGTGTGAACAGCGCACCGGCTCGGCGCCAGCTCTCCCAACAGCTCCCAACAGC 421
Db 236 SerValCysAsnSerAlaProGlySerGlyProSerSerProAsnSerSerHisSerThr 255
QY 422 ATCGTGAGAAATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCAG 481
Db 256 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGln 275
QY 482 CACGAGCCCTCTCTGGACAGTCCCAACAGCTTACAGCTCTACAGCTCTCCTCTCT 541
Db 276 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTyThrSerProSer 295
QY 542 CTGCCCCAACATCTCCCTAGGCTGCAGCGCCAGGCTCACTGTCAACCACTCACACCTCACT 601
Db 296 LeuProAsnIleSerLeuGlyLeuGlnAlaThrValThrValThrAsnSerHisLeuThr 315
QY 602 GCCTCCCGGAGCTGTGCACACAGCAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 661
Db 316 AlaSerProLysLeuSerThrGlnGlnGluAlaGluArgGlnAlaLeuGlnSerLeuArg 335
QY 662 CAGGCTGGCAGCTGACCGCAAGTTTATGAGCACATCTCTATTCTGCTGCTGCTGCTG 721
Db 336 GlnGlyGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCysLeuLeu 355
QY 722 GCGGTGGCACTGGAG 736
Db 356 GlyValAlaLeuGlu 360

RESULT 9
Q91X14
```


Db	354	----	-----GlyGluGlyLeuAlaGluLeuTyr	361
QY	1328	TCAGTGCCTCCCAACCGCTGCAACCTTTTCAGAGTGTACCAAGAGCGCCCTCAGCCTGGCCACT	1387	
Db	362	AsnGlyLeuLeuProGlnAlaLysLeuGlnLeuGlnAla-----	375	
QY	1388	GTGCCCCACCAAGCCTCGGCCGTACCCANCTCCCTGCTGCCCTCGGGGCGATG----	1444	
Db	376	-----LeuAlaAlaGluSerGlyPheLeu	383	
QY	1445	---AAGAACCCCCACACCAACCGCTCAAGCACCTCTTCACCACCAAGTGTGGTCTACGAC	1501	
Db	384	AlaLysGlnGluPro-----ThrCysThrThrGlyLeuGlyTyrAsp	397	
QY	1502	ACGTTTCATGCTAAAGCACCAAGTGCATGTGCGGGAACACACACGCTGCACCTCGAGCAATGCT	1561	
Db	398	GlnAlaMetValArgHisGluCysCysGlyAsnAsnAlaSerHisValGluAsnGly	417	
QY	1562	GGCGGATCCAGACATCTGGTCCCGGCTGCAGGAGACAGCGCTGCTTAGCAAGTGGCAG	1624	
Db	418	GlyArgIleGlnSerIleTrpSerLysLeuIleGluHisGlyHisValGlnLysCysGlu	437	
QY	1622	CGATCCGAGGTGCGAAAGCCAGCTAGATGATGATCCAGACAGTGCCTCTGAATACCAC	1681	
Db	438	LysValThrAlaLysLysAlaSerLeuGluGlnLeuValHisSerGlnThrTyr	457	
QY	1682	ACCTGCTCTATGGGACAGTCCCTCTCAACCGCGAGAAGCTAGACAGCAAGTTCCTC	1741	
Db	458	ThrThrPhePheAlaValSerProThrAlaCysLeuLysIleAspAlaAsnSerLeu---	476	
QY	1742	GGTCCCATCAGCAAGATGATGCTGTGCTGCTTGTGGGGGATGTCGGGTGGACAGT	1801	
Db	477	---ProLeu-----LysArgPheLeuGlnLeuProCysGlyIleGlyValAspSer	493	
QY	1802	GACACCGTGTGGAATCAGATGCACCTCTCCAGCTGCTGCCATCGGACGTGGGTGGCTG	1861	
Db	494	AspThrTyrPheAsnAspAlaSerThrGlnThrAlaAlaArgLeuAlaAlaGlyThrLeu	513	
QY	1862	CTGGAGCTGGCCCTCAAGTGGCTGCAGGAGAGCTCAAGAATGGATTTGCCATCATCCGG	1921	
Db	514	IleGluLeuSerSerGlnValAlaGluGlyArgLeuLysAsnGlyPheAlaCysIleArg	533	
QY	1922	CCCCAGGACACACCGCGGAGNATCCACAGCCATGGGATCTGCTTCTTCAACTCTGTGA	1981	
Db	534	ProGlyHisHisAlaGluHisGluGlnAlaMetGlyPheCysPhePheAsnAsnVal	553	
QY	1982	GCCATCACCCCAAACTCTACAGCAGAG-----TTGACGTGGGCAAGTCTCTCATC	2035	
Db	554	AlaValAlaValLysValLeuGlnThrTyrProAlaGlnCysAlaLysIleAlaIle	573	
QY	2036	GTGACGTGGGACATTCACCATGGCAATGCACCCACGAGCGCTTCACAAATGACCCCTCT	2095	
Db	574	IleAspTrpAspValHisHisGlyAsnGlyThrGlnLeuSerPheGluAsnAspProAsn	593	
QY	2096	GTGCTCTACATCTCTGTCATCGCTATGACAACGGGAACCTCTTTCCAGGCTCTGGGCT	2155	
Db	594	ValLeuTyrMetSerLeuHisArgHisAspLysGlyAsnPhePheProGlyThrGlySer	613	
QY	2156	CCTGAAGAGCTGTGTGGAGGACACGCGCTGGGTACATGCAAGCTGGCATGGACAGGA	2215	
Db	614	ValThrGluValGlyLysAsnAspAlaLysGlyLeuThrValAsnValProPheSerGly	633	
QY	2216	GGTGTGACCCCCCATTTGGAGACGTGGAGTAGTACCTTACAGCCTTCAGCAGAGTGGTGATG	2275	
Db	634	AspVal-----MetArgAspProGluTyrLeuAlaAlaTrpArgThrValIleGlu	650	
QY	2276	CCCATTGCCACGAGTTCTTCACATGATGGTCCCTAGTCTCCGCGGGTTTGTATGCTGTT	2335	
Db	651	ProValMetAlaSerPheCysProAspPheIleIleValSerAlaGlyPheAspAlaCys	670	
QY	2336	GAAGACATCTGTCTCTCTGGTGCTACTCTGTACCGCGCCAGATGTTTGGCCACTWTG	2395	
Db	671	HisGlyHisProAsnAlaLeuGlyGlyTyrGluValThrProGluMetPheGlyTyrMet	690	

Qy	2396	ACCAGCAGCGTGTATGACCCCTGCAGCGGGGCCGGGNGTGCTGGCCCTGGAGGAGGCCAT	2435
Df	691	ThrlSerSerLeuLeuAsnTrAlaSerGlyLysValValLeuAlaLeuGlugLytyr	710
Qy	2456	GACTTCAGCGCATCTGTCATCGCTCTGAAGCTTGCTGCTCGGCTCTGCTC-----AGT	2509
Df	711	AspleuLysSerIleSerGluAlaAlaGlnGlnCysValGlnAlaLeuIleGlyLysUser	730
Qy	2510	GTAAGCTCGACGCCCTTTGATGAGGCACTTTTGCAGAAAAAGCCCCAACATCAACGCACTG	2569
Df	731	AspAspAlaGlyArgLeuSerSerValAlaLeuGluSerLeuProAsnProSerAlaVal	750
Qy	2570	GCCACGCTAGAGAAA	2584
Df	751	GLuThrLeuGlnLys	755
RESULT 11			
ID	Q96K29	PRELIMINARY;	PRT; 191 AA.
AC	Q96K29;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DE	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	CNA FLJ14828 fis, clone OVARC1000915, highly similar to Homo sapiens histone deacetylase 5 mRNA.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxId=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=OVARIAN CARCINOMA;		
RA	Isoqai T., Ota T., Hayashi K., Sugiyama T., Olsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,		
RA	Wagatsuma M., Hosoiri T., Raku Y., Kondra H., Kondo H., Sugawara M.,		
RA	Takanashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,		
RA	Yamanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,		
RA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,		
RA	Ninomiya K., Iwayanagi T.,		
RT	"NEDO human cDNA sequencing project";		
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBDJ databases.		
DR	EMBL; AK027734; BAB55329.1; --		
DR	Interpro; IPR000286; His_deacetylase.		
DR	Pfam; PF00850; Hist_deactyl; 1.		
SQ	SEQUENCE 191 AA; 20877 MW; DCB6D45C0756A988 CRC64;		
Alignment Scores:			
Pred. No.:	4,74e-49	Length:	191
Score:	945.00	Matches:	179
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	17.68%	Indels:	0
DB:	4	Gaps:	0
US-09-502-945-2 (1-2885) x Q96K29 (1-191)			
Qy	1508	ATGCTAAAGCACCGTAGTCATGTGCGGGAGACACACACAGCTGCACCTGTAGCATGCTGGCGG	1567
Df	1	MetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGluHisAlaGlyArg	20
Qy	1568	ATCCAGAGCATCTGGTCCCGCTGCAGGAGACAGCGCTCTTAGCAAGTGGCAGCGGATC	1627
Df	21	IleGlnSerIleIrrpSerArgLeuGlnGluThrGlyLeuLeuSerLysCysGluArgIle	40
Qy	1628	CGAGGTCGCAAAGCCAGCTAGATGATGATCCAGACAGTGCATCTTGAATACACACCCCTG	1687
Df	41	ArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGluTy-HisThrLeu	60
Qy	1688	CTCTATGGGACCATGCTCCCTCAACCGGACAGAGCTAGACAGACAAGAAGTTCTCGTCCC	1747
Df	61	LeutvrgLysThrSerProLeuAsnArqGlnLysLeuAspSerLysLysLeuGlyPro	80

QY	1748	ATCAGCCAGAGATGTATGCTGCTGCTGGGGCATCGGGTGGACAGTGCACACC	1807
Db	81	IleSerGlnLysMetTyrAlaValLeuProCysGlyGlyIleGlyValAspSerAspThr	100
QY	1808	GTGTGGAGATGAGTCACTCTCCAGTCTGTGCGCATGGCAGTGGGCTGCTCGTGGAG	1867
Db	101	ValTrpAsnGluMetHisSerSerAlaValArgMetAlaValGlyCysLeuLeuGlu	120
QY	1868	CTGGCCTTCAAGTGGCTCGAGAGAGCTCAAGAATGGATTGGCATCATCCGGCCCCCA	1927
Db	121	LeuAlaPheLysValAlaAlaGlyGluLeuLysAsnGlyPheAlaIleArgProPro	140
QY	1928	GGACACACCGCCGAGGAATCCACAGCCATGGGATCTCTCTTCAACTCTGTAGCCATC	1987
Db	141	GlyHisAlaGluGluSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIle	160
QY	1988	ACCCGAAAACCTCTACAGCAGAAAGTGAAGCTGGCGAAGGTCTCATCTGGAGCTGG	2044
Db	161	ThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeuIleValAspTrp	179
RESULT 12			
Q9NWA9			
ID	Q9NWA9	PRELIMINARY;	PRT; 276 AA.
AC			
DT	01-OCT-2000	(TREMBlrel. 15, Created)	
DT	01-OCT-2000	(TREMBlrel. 15, Last sequence update)	
DT	01-JUN-2002	(TREMBlrel. 21, Last annotation update)	
DE	Hypothetical 29.6 kDa protein.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=EMBRYO;		
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,		
RA	Wagatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,		
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,		
RA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masubo Y.,		
RA	Ninomiya K., Iwayanagi T.		
RT	"NEDO human cDNA sequencing project."		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AK001032; BAA91474.1; -		
DR	InterPro: IPR000286; His_deacetylase.		
DR	Pfam: PF00850; Hist_deacetyl; 1.		
DR	PRINTS; PR01270; HDASUPER.		
SQ	SEQUENCE 276 AA; 29622 MW; 8C96C7C5B41CD3A1 CRC64;		
Alignment Scores:			
Pred. No.:	5.38e-47	Length:	276
Score:	911.50	Matches:	178
Percent Similarity:	75.53%	Conservative:	35
Best Local Similarity:	63.12%	Mismatches:	56
Query Match:	17.05%	Indels:	14
DB:	4	Gaps:	2
US-09-502-945-2 (1-2885) x Q9NWA9 (1-276)			
QY	1955	ATGGGATCTGCTTCTCACTCTGTAGCCATCCCGCAAACTCTACAGCAGAGTTG	2014
Db	1	MetGlyPheCysPhePheAsnSerValAlaIleAlaCysArgGlnLeuGlnGlnSer	20
QY	2015	AACGTGGGCAAGGTCTCATCTGGAGTGGGACATTCACCATGGCAATGGCAGCCAGCAG	2074
Db	21	LysAlaSerLysIleLeuIleValAspTrpAspValHisGlyAsnGlyThrGlnGln	40
QY	2075	GGTCTTCAAGTACCCCTCTGTGCTCTACATCTCTCTGCTATGACACGGGAG	2134
Db	41	ThrPheTyrGlnAspProSerValLeuTyrIleSerLeuHisArgHisAspGlyAsn	60
QY	2135	TTCCTTCCAGGCTCTGGGCTCTCTGNAGAGGTTGGTGGAGGACCGCTGGGTACAA	2194

Db	61	PhePheProGlySerGlyAlaValAspGluValGlyAlaGlySerGlyGluGlyPheAsn	80
QY	2195	GTGAACGTGGCATGGACAGGAGTGTGGACCCCATTCATGGAGCTGGAGTACCTTACA	2254
Db	81	ValAsnValAlaTrpAlaGlyGlyLeuAspProMetGlyAspProGluTyrLeuAla	100
QY	2255	GCCTTCAGGACAGTGTGTATGCCCATTCGCCACGAGTCTTCACCTGATGTGGTCTAGTC	2314
Db	101	AlaPheArgIleValValMetProIleAlaArgGluPheSerProAspLeuValLeuVal	120
QY	2315	TCCGCGGGTTCATGCTGTGTGAAGGACATCTGCTCTCTCTGGGTGGCTACTCTGTACCC	2374
Db	121	SerAlaGlyPheAspAlaAlaGlyHisProAlaProLeuGlyGlyTyrHisValSer	140
QY	2375	GCCAGATGTTTTCGCCACCTTGACACAGGAGCTGATGACCTTGGCAGGGGCCGGTGGTG	2434
Db	141	AlaLysCysPheGlyTyrMetThrHisGlnLeuMetAsnLeuAlaGlyGlyAlaValVal	160
QY	2435	CTGGCCCTGGAGGAGGCCATGACTTGACCGCCATCTGTGATGCCCTCTGAAGCTTGTGTC	2494
Db	161	LeuAlaLeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysVal	180
QY	2495	TCGGCTCTGCTCAGTGTAAAGCTGCAGCCCTTGGATGGAGGAGCTCTTGCAGCAAGCCCC	2554
Db	181	AlaAlaLeuLeuGlyAsnArgValAspProLeuSerGluGluGlyTrpLysGlnLysPro	200
QY	2555	ACATCAACGACGTGGCCAGCTAGAGAAAGTCAATCAGATCCAGAGCAAACTCGGAGC	2614
Db	201	AsnLeuAsnAlaIleArgSerLeuGluAlaValIleArgValHisSerLysTyrTrpGly	220
QY	2615	TGTGTGCAGAAAGTTCGCGCGTGGTCTGGCGCGTCCCTGCGAGGGGCCCAAGCAGCTGAG	2674
Db	221	CysMetGlnArgLeuAlaSerCysProAspSerTrpValProArgValProGlyAlaAsp	240
QY	2675	ACCGAAGAACCGCAAT-GTGAACGCCATGGCTGCTGCTGTGGTGGGG-----GCCGAA	2727
Db	241	LysGluGluValGluAlaValThrAlaLeuAlaSerLeuSerValGlyIleLeuAlaGlu	260
QY	2728	CAGGCCCAAGCTGCGGCAGCCCGGGAACACAGCCCCCGCAGAGAGCCCATGGAG	2787
Db	261	Asp-----ArgProSerGluGlnLeuValGlu	269
QY	2788	CAGGAG 2793	
Db	270	GluGlu 271	
RESULT 13			
Q9YGY4			
ID	Q9YGY4	PRELIMINARY;	PRT; 565 AA.
AC			
DT	01-MAY-1999	(TREMBlrel. 10, Created)	
DT	01-MAY-1999	(TREMBlrel. 10, Last sequence update)	
DT	01-DEC-2001	(TREMBlrel. 19, Last annotation update)	
DE	MITR protein.		
GN	MITR.		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8355;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=95417589; PubMed=10487760;		
RA	Sparrow D.B., Miska E.A., Langley E., Reynaud-Deonauth S., Kotecha S.,		
RA	Towers N., Spohr G., Kouzarides T., Mohun T.J.;		
RT	"MEF-2 function is modified by a novel co-repressor, MITR.";		
RL	EMBO J. 18:5085-5098(1999).		
DR	EMBL: Z97214; CAB10167.1; -		
SQ	SEQUENCE 565 AA; 63777 MW; 29B778233589F1CD CRC64;		
Alignment Scores:			
Pred. No.:	3.91e-42	Length:	565


```

QY 1072 CACCACCCACCTGAGGAGACAGAGGAGCTGACGGAGCAGCAGGAGCTGCTGGG 1131
Db 219 His .....-LeuLeuMetGlyLeuAlaGly 226
QY 1132 GGAGGGAGCCCTGACCATGCCCGGAGGAGCTCCACAGAGAGTGAGACACAGGAAGA 1191
Db 227 GlyLeuLeuLeuSerLeuGluGlyGlyTyraAsnLeu ..... 239
QY 1192 CCTGGAGGAGGAGCAGGAGGAGGATGGGAGGAGGAGGATTCAGCTCAGGTTAA 1251
Db 240 .....-ArgAlaLeuAlaGluGlyValSerAlaSerLeuHis-ThrLeuLe 254
QY 1252 GGAGGAGGAGGAGGAGAGTGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1311
Db 254 u- .....-GlyAspProGlyProMetLeuGluSerProGlyAlaPr 267
QY 1312 ATACAAAAAAGCTGCTCAGATGCCAA---CCGCTCAACCTTTGAGGTGTACCAAGC 1368
Db 267 ocysArgSerAlaGlnAlaSerValSerCysAlaLeuGluAlaLeuGluProhetrpgl 287
QY 1369 GCCCTC-----AGCTGGCCACTGTGCCCCCAGGAGGAGGAGGAGGAGGAGGAGGAG 1422
Db 287 uValLeuValArgSerThrGluThrValGluArgGlyHisMetGlyGluAspAsnValGl 307
QY 1423 CCCTGCTGCCCTGGGGGAGTGAAGAACCCCGAGACCAACCGCTCAAGCAC----- 1474
Db 307 uGluSerGluGluGluGlyProtrpgluProProValLeuProIleuThrTrpProVa 327
QY 1475 -----CTCTTCCACCAAGTGTGTACGACAGCTTATCTGCTAAAGCAGCAGTGCTGTG 1530
Db 327 lLeuThrSerArgThrGlyLeuValTyraAspGlnAsnMetMetAsnHis---CysAsnLe 346
QY 1531 CGGGAACACACAGTGCACCTGAGCATGCTGCCGAGATCCAGAGCATGTGTCGGGCT 1590
Db 346 utrpAspSerHis---HisProGluValProGlnArgIleLeuArgIleMetCysArgLe 365
QY 1591 GCAGGAGACAGGCTGCTAGCAAGTGCAGGAGTGCAGAGTGCAGAAAGCCACGCTAGA 1650
Db 365 uGluGluLeuGlyLeuAlaGlyArgCysLeuThrLeuThrProArgProAlaThrGluAl 385
QY 1651 TGAGATCCAGACAGTGCATCT---GAATACACACCTGCTCTATGGACCACTGCCCT 1707
Db 385 aGluLeuLeuThrCysHisSerAlaGluTyraValGlyHisLeuArgAlaThrGluLysMe 405
QY 1708 CAACCGGACAGACAGCAGCAAGAGTGTGCTGCCATCCATCAGCAGAGATGATGC 1767
Db 405 tlyThrArgGluLeuHisArgGlu----- 413
QY 1768 TGTGCTGCCTTGTGGGGATCGGGTGGAGTGCACCGTGTGGAATGAGATGCACATC 1827
Db 414 -----SerSerAsnPheAspSerIleTyriLeCysProSerTh 426
QY 1828 CTCCAGTGTGTGCGCATGGCAGTGGGCTGCCCTGGAGCTGGCCTTCAAGTGGCTGC 1887
Db 426 rPheAlaCysAlaGlnLeuAlaThrGlyAlaAlaCysArgLeuValGluAlaValLeuSe 446
QY 1888 AGGAGAGCTCAAGATGGATTTCCATCATTCGGCCCCCAGAGACACCCAGCGGAGGAATC 1947
Db 446 rGlyGluValLeuAsnGlyAlaAlaValValArgProProGlyHisHisAlaGluGlnAs 466
QY 1948 CACAGCATGGGATTCGCTCTTCAACTCTGTAGCCATCCAGCAAACTCCTACAGCA 2007
Db 466 pAlaAlaCysGlyPheCysPhePheAsnSerValAlaValAlaAlaAlaGlnHisAlaGlnTh 486
QY 2008 G---ANGTTGAGCTGGGCAAGTCTCATCTGAGTGGGACTGGGACATTCACCATGGCAATGG 2064
Db 486 rIleSerGlyHisAlaLeuAlaArgIleLeuIleValAspTrpAspValHisHisGlyAsnGl 506
QY 2065 CACCAGCAGGCTTCAATGATGACCCCTGCTGTGCTCTACATCTCTCTCATCGCTATGA 2124
Db 506 yThrGlnHisMetPheGluAspAspProSerValLeuTyraValSerLeuHisArgTyraS 526
QY 2125 CAACGGGAATCTCTTTCCA-----GGCTCTGGGCTCTCTGAAGAGTGTGGTGGAGGACC 2178

```

```

RESULT 15
Q9NSW6
ID Q9NSW6 PRELIMINARY; PRT; 1063 AA.
AC Q9NSW6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 114.4 kDa protein.
GN DKFZP566E044.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Bloeker H., Boecher M., Brandt P., Mewes H.W., Gassenhuber J.,
RA Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137696; CAB70878.1;
DR InterPro; IPR000286; His_deacetylase.
DR InterPro; IPR001607; ZnF_UBP.

```

```

Db 526 pHisGlyThrPhePheProMetGlyAspGluGlyAlaSerSerGlnIleGlyArgAlaAl 546
QY 2179 AGCGTGGGTACAATGTGAACCTGGCATGGCAGGAGGTGTGGACCCGCCCATTTGGAGA 2238
Db 546 aGlyThrGlyPheThrValAsnValAlaTrpAsnGly-----ProArgMetGlyAs 563
QY 2239 CGTGGAGTACCTTACAGCCTTCAGGACAGTGTGGTATGCCCATTTGCCAGAGTTCACCC 2298
Db 563 pAlaAspTyrrLeuAlaAlaTrpHisArgLeuValLeuProIleAlaTyrrGluPheAsnPr 583
QY 2299 TGATGTGGTCTAGTCTCCCGGGTTTGTATGCTGTGTGAAGACATCTGCTCTCTCTGGG 2358
Db 583 oGluLeuValLeuValSerAlaGlyPheAspAlaAlaArgGly-----AspProLeuGl 601
QY 2359 TGCTACTCTGTCCACCGCCAGATGTTTGGCCACTTGAACGAGCAGCTGATGACCCCTGGC 2418
Db 601 yGlyCysGlnValSerProGluGlyTyrrAlaHisLeuThrHisLeuLeuMetGlyLeuAl 621
QY 2419 AGGGGCCGGGTGTGCTGCGCCCTGGAGGAGGCCATGACTTGACCGCCATCTGTGATGC 2478
Db 621 aSerGlyArgIleIleLeuIleLeuGluGlyGlyTyrrAsnLeuThrSerIleSerGluSe 641
QY 2479 CTCTGAAGCTTGTCTCGGCTCTGCTCAAGTGTAAAGCTGCAGCCCTTGGATGAGGCACT 2538
Db 641 rMetAlaAlaCysThrArgSerLeuLeuGly-----AspProProProLeuLeuTh 658
QY 2539 CTTCGAGCAAAAGCCCAACATCAACGACCTGGCCGCTAGAGAAAGTATCATGAGATCCA 2598
Db 658 rLeuProArgProProLeuSerGlyAlaLeuAlaSerIleThrGluThrIleGlnValHi 678
QY 2599 GAGCAAACTGAGAGTGTGTGCAGAAAGTTCGCGCTGCTGGCGGCTCCCTGCGAGG 2658
Db 678 sAlaArgTyrrCysSerLeu-Arg-----ValMetLysValGluAspArgGluG 695
QY 2659 GCGCC-----AAGCAGGTGAGACCGGAGAGAGCCGAAATGT- 2693
Db 695 lyProSerSerSerLysLeuValThrLysLysAlaProGlnProAlaLysProArgLeuA 715
QY 2694 -----GAACGCCATGGCTTGTGTTGGTGGGCGCGA---ACAG 2730
Db 715 laGluArgMetThrThrArgGluLysLysValLeuGluAlaGlyMetGlyLysValThrS 735
QY 2731 GCCCAAGTGGCGG-----AGCCCGGGA----- 2753
Db 735 eAlaSerPheGlyGluGluSerThrProGlyGlnThrAsnSerGluThrAlaValAla 755
QY 2754 -----ACACAGCCCGCAGCGCGCAGGAGCCCATGGACGAGGAGCTGCCCTGTGACG 2807
Db 755 laLeuThrGlnAspGlnProSerGluAlaAlaThrGlyGlyAlaThrLeuAlaGlnThr 774

```


DR Pfam: PF00850; Hist_deacetyl1; 2.
DR Pfam: PF02148; zf-UBP; 1.
DR PRINTS: PR01270; HDASUPER.
DR SMART: SM00290; ZnF_UBP; 1.
KW Hypothetical protein.
SQ SEQUENCE 1063 AA; 114422 MW; C0040EBD4355F62C CRC64;

Alignment Scores:
Pred. No.: 4,98e-39 Length: 1063
Score: 781.50 Matches: 200
Percent Similarity: 50.91% Conservative: 81
Best Local Similarity: 36.23% Mismatches: 191
Query Match: 14.62% Indels: 80
DB: 4 Gaps: 19

US-09-502-945-2 (1-2885) x Q9NSW6 (1-1063)

QY 1286 CCCGACTTGGAGGAGCTGTGCTGGATACAAAAAAGTGTCTCAGATGCCCAA---CCG 1342
Db 256 ProMetLeuGluSerProGlyAlaProCysArgSerAlaGlnThrSerValSerCysAla 275
QY 1343 CTGCAACCTTTGAGGTGTACCAAGCCCTC-----AGCCTGGCCACTGTGCCCCAC 1396
Db 276 LeuGluAlaLeuGluProPheTrpGluValLeuValArgSerThrGluThrValGluArg 295
QY 1397 CAAGCCCTGGCCGATACCAATCTCCCTGCTGCCCTGGGGCATGAAGAACCCCCCA 1456
Db 296 AspAsnMetGluGluAspAsnValGluGluSerGluGluGluGlyProTrpGluProPro 315
QY 1457 GACCAACCCGTCAAGCACTCTTCACG-----ACAAGTGTGGTC 1495
Db 316 ValLeuProLeu-----LeuThrTrpProValLeuGlnSerArgThrGlyLeuVal 332
QY 1496 TAGGACAGCTTCATGCTAAAGCACCAGTGTGCGGGAAACACACAGTGCACCTGAG 1555
Db 333 TyrAspGlnAsnMetMetAsnHis---CysAsnLeuTrpAspSerHis---HisProGlu 350
QY 1556 CATGCTGGCCGGATCCAGACATCTGTCCCGCTGCAGGAGACAGCCCTGCTTAGCAAG 1615
Db 351 ValProGlnArgIleLeuArgIleMetCysArgLeuGluGluLeuGlyLeuAlaGlyArg 370
QY 1616 TGGAGCGGATCGAGGTCCGAAGCCAGCTAGTGTAGATCCAGACAGTGCCTCT--- 1672
Db 371 CysLeuThrLeuThrProArgProAlaThrGluAlaGluLeuLeuThrCysHisSerAla 390
QY 1673 GAATACACACCTGCTCTATGGACACAGTCCCTCAACCCGAGAGCTAGACAGCAAG 1732
Db 391 GluTrpValGlyHisLeuArgAlaThrGluLysMetLysThrArgGluLeuHisArgGlu 410
QY 1733 AAGTTGCTCGGTCCCATCAGCCAGAGATGTATGCTGTGCTGTGGGGGCATCGGG 1792
Db 411 -----Ser 411
QY 1793 GTGACAGTGCACACCGTGTGGATGAGATGCATCTCCAGTGTGTCGATGGCGAGTG 1852
Db 412 SerAsnPheAspSerIleTyrIleCysProSerThrPheAlaCysAlaGlnLeuAlaThr 431
QY 1853 GGCTGCTGTGGAGTGGCTTCAAGTGGCTCAGGAGAGCTCAGATGGATTGCTTGC 1912
Db 432 GlyAlaAlaCysArgLeuValGluAlaValLeuSerGlyGluValLeuAsnGlyAlaAla 451
QY 1913 ATCATCCGGCCCGGACACACCGCCGAGGAATCCACAGCTGGGATTTCTGCTTCTTC 1972
Db 452 ValValArgProProGlyHisHisAlaGluGlnAspAlaCysGlyPheCysPhePhe 471
QY 1973 AACTCTAGCTAGCTACCCGCAAACTCTACAGCAG---AAGTTGAAGTGGGCAAGTGC 2029
Db 472 AsnSerValAlaValAlaAlaArgHisAlaGlnThrIleSerGlyHisAlaLeuArgIle 491
QY 2030 CTCATCTGGACTGGGACATTCACCATGGCAATGGCCAGCCAGAGCGGTCTTCAATGAC 2089
Db 492 LeuIleValAspTrpAspValHisHisGlyAsnGlyThrGlnHisMetPheGluAspAsp 511

QY 2090 CCTCTGTGCTCTACATCTCTCATGATGACACGGGAACCTTCTTCCA----- 2143
Db 512 ProSerValLeuTyrValSerLeuHisArgTyrAspHisGlyThrPhePheProMetGly 531
QY 2144 GCCTCTGGGCTCCTGAAGAGCTTGGTGGAGGACCCAGCGGTGGGTCAAACTGTAACCTG 2203
Db 532 AspGluGlyAlaSerSerGlnIleGlyArgAlaAlaGlyThrGlyPheThrValAsnVal 551
QY 2204 GCATGGACAGAGAGTGTGGACCCCATTTGGAGAGCTGGAGTACCTTACAGCCTTACAG 2263
Db 552 AlaTrpAsnGly-----ProArgMetGlyAspAlaAspTyrLeuAlaAlaTrpHis 568
QY 2264 ACAGTGGTATGCCCATTTGCCACAGTCTCACCCTGATGTGGTCTCTAGTCTCCCGCGG 2323
Db 569 ArgLeuValLeuProIleAlaTyrGluPheAsnProGluLeuValLeuValSerAlaGly 588
QY 2324 TTTGATGCTGTTGAAGGACATCTCTCTGCTGGTGGTACTCTGTCCACGCCAGATGT 2383
Db 589 PheAspAlaAlaArgGly-----AspProLeuGlyGlyCysGlnValSerProGluGly 606
QY 2384 TTTGGCCACTTGACAGGACGCTGATACCTTGGCGAGGGCCGGGTGGTGTGTCGCCCTG 2443
Db 607 TyrAlaHisLeuThrHisLeuLeuMetGlyLeuAlaSerGlyArgIleIleLeuIleLeu 626
QY 2444 GAGGAGGCGCATGACTTGACCCCATCTGTGATGCTCTGAAGCTTGTCTCGGCTCTG 2503
Db 627 GluGlyGlyTyrAsnLeuThrSerIleSerGluSerMetAlaAlaCysThrArgSerLeu 646
QY 2504 CTCAGTGTAAAGCTGCAGCCCTTGGATGAGGCGAGTCTTGACAGAAAAGCCCAACATCAAC 2563
Db 647 LeuGly-----AspProProLeuLeuThrLeuProArgProProLeuSerGly 663
QY 2564 GCATGGCCACCTAGAGAAAGTCTCGAGATCCAGAGCAAAACACTGGAGCTGTGTGCAG 2623
Db 664 AlaLeuAlaSerIleThrGluThrIleGlnValHisArgArgTyrTrpArgCysLeu-Ar 683
QY 2624 AAGTTCGCGCTGTGCTGGCGCGTCCCTCGAGGGGCGCC----- 2663
Db 683 g-----ValMetLysValGluAspArgGluGlyProSerSerSerLysLeuAspTh 700
QY 2664 ----AAGCAGGTGAGACCCGAAAGCCGAATGT-----GAAGC 2698
Db 700 rLysLysAlaProGlnProAlaLysProArgLeuAlaGluArgMetThrThrArgGlu 720
QY 2699 CCATGGCCTTGTGTGGTGGGGCGCA---ACAGCCCAAGCTGCGGC-----AG 2746
Db 720 sLysValLeuGluAlaGlyMetGlyLysValThrSerAlaSerPheGlyGluGluSerTh 740
QY 2747 CCGCGGA-----ACACAGCCCGCGCGCGCAG 2773
Db 740 rProGlyGlnThrAsnSerGluThrAlaValValAlaLeuThrGlnAspGlnProSerG 760
QY 2774 AGGAGCCCTGGAGCAGGAGCCCTGCCCTGTGAGC 2807
Db 760 uAlaAlaThrGlyGlyAlaThrLeuAlaGlnThr 771

Search completed: March 21, 2003, 12:58:36
Job time : 136.938 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 08:38:27 ; Search time 31.7038 Seconds
(without alignments)
10910.938 Million cell updates/sec

Title: US-09-502-945-3

Perfect score: 2140

Sequence: 1 ggctgctgaatgactgcga.....gaagaagaagaaggaaaaa 1298

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09502945/runat_14032003_101057_19087/app-query.fasta_1.10979
-DB=A_Geneseq_101002 -QWTF=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945_ECGN_1.1_365_runat_14032003_101057_19087 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1047	48.9	236	20	Colon cancer assoc
2	677	31.6	157	22	Human protein sequ
3	401	18.7	434	18	Antibiotic potenti
4	401	18.7	434	21	(Lys)434 protein s
5	365	17.1	554	22	Drosophila melanog
6	354.5	16.6	808	22	Novel human diagno
7	326	15.2	71	21	Human secreted pro
8	325.5	15.2	215	22	Novel human diagno
9	325.5	15.2	215	22	Novel human diagno
10	314.5	14.7	182	22	Novel human diagno
11	313	14.6	611	20	T. gondii immunoge
12	313	14.6	611	22	T. gondii immunoge
13	312	14.6	274	22	Drosophila melanog
14	309.5	14.5	771	22	Drosophila melanog
15	309	14.4	405	22	Novel human diagno
16	307.5	14.4	1178	18	Novel human diagno
17	304	14.2	436	22	Mannose-1-phosphat
18	298	13.9	1898	20	Novel human diagno
19	290.5	13.6	389	22	A human trichohyal
20	286.5	13.4	328	22	Novel human diagno
21	285.5	13.3	573	22	Novel human diagno
22	281	13.1	843	22	Human protein SEQ
23	280	13.1	204	22	Novel human diagno
24	277.5	13.0	191	22	Novel human diagno
25	276.5	12.9	695	22	Novel human diagno
26	273	12.8	189	22	Human protein SEQ
27	266.5	12.5	347	22	Novel human diagno
28	266	12.4	140	22	Human polypeptide
29	266	12.4	624	23	Novel human diagno
30	265	12.4	100	21	Human RNA metaboli
31	265	12.4	100	21	Peptide #8 used in
32	265	12.4	100	22	Amino acid polymer
33	265	12.4	100	22	Poly-L-Lysine used
34	264	12.3	146	22	Nucleic acid trans
35	264	12.3	1098	22	Novel human diagno
36	264	12.3	1333	22	Novel human diagno
37	264	12.3	1430	22	Drosophila melanog
38	263	12.3	361	20	Novel human diagno
39	261.5	12.2	194	22	Biorhythm marker p
40	261.5	12.2	540	22	Novel human diagno
41	261	12.2	1074	22	Human Ser/Arg rich
42	261	12.2	1507	21	Novel human diagno
43	259	12.1	412	17	Plasmodium yoelii
44	259	12.1	735	22	Human thyrotropin
45	258	12.1	191	22	Novel human diagno

ALIGNMENTS

RESULT 1
AA07093

ID RAY07093 standard; Protein; 236 AA.

XX RAY07093;

XX 02-JUL-1999 (first entry)

XX Colon cancer associated antigen precursor sequence.

XX Cancer associated antigen; diagnosis; research; treatment; human;
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX prostate cancer.

OS Homo sapiens.

PN WO9904265-A2.

XX 28-JAN-1999.

PD

DR N-PSDB; AAH99850.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
PS Claim 20; Page 289; 1217pp; English.
XX
XX
CC AAH99166 to AAH99904 encode the human proteins given in AAH25225 to
CC AAH25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: anti-inflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
CC antitumor; osteopathic; dermatological; antiallergic; antisthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
SQ Sequence 157 AA;
Alignment Scores:
Pred. No.: 2,44e-56 Length: 157
Score: 677.00 Matches: 150
Percent Similarity: 95.54% Conservative: 0
Best Local Similarity: 95.54% Mismatches: 7
Query Match: 31.64% Indels: 2
DB: 22 Gaps: 0
US-09-502-945-3 (1-1298) x AAH25909 (1-157)
QY 11 ATGACTGCGAACCGCTTGCAGAGACCTTCTGGCTTGAGCCANCAAGAGAACTAGCG 70
Db 1 MetThrAlaAsnArgLeuAlaGluSerLeuLeuAlaLeuSerGlnGlnGluLeuAla 20
QY 71 GATTTGCCAAAGACTACTCTTGAGTGAGAGTGAAGATGAGGGGACAATGATGAGAG 130
Db 21 AspLeuProLysAspTyrLeuLeuSerGluSerGluAspGluGlyAspAsnAspGlyGlu 40
QY 131 AGAAAGCATCNAAGCTTCTGGAACCAATCAGTTCCTCCCTTGATGGAAGAANTAGCGGAAA 190
Db 41 ArgLysHisGlnLysLeuLeuGluAlaIleSerSerLeuAspGlyLysAsnArgArgLys 60
QY 191 TTGGCTGANAGGCTCTGAGGCTAGTCTGAAGGTGTGAGAGTTCAGTTCAGTTCCTGAAGGA 250
Db 61 LeuAlaGluArgSerGluAlaSerLeuLysValSerGluPheAsnValSerSerGluGly 80
QY 251 TCAGGAGAAAAGCTGGTCCCTTGCAGATCTGCTTGAGCCTGTAAACCTTCATCTCTTTG 310
Db 81 SerGlyGluLysLeuValLeuAlaAspLeuLeuGluProValLysThrSerSerLeu 100
QY 311 GCCACTGTGAAAAGCAACTGAGTAGAGTCANATCAAGAANACAGTGGAGTTACCTCTG 370
Db 101 AlaThrValLysLysGlnLeuSerArgValLysSerLysLysThrValGluLeuProLeu 120
QY 371 AACAAAGAGAGATTGAACGATCCACAGAGAATA-GCATTCATAAACAACCA-CAGATC 428
Db 121 AsnLysGluGluIleGluArgIleHisArgGluValAlaPheAsnLysThrAlaGlnVal 140
QY 429 CTCTCCAAATGGGACCTGTGCTCTGAAGAACCGGACGAGGAGCAGACTG 479
Db 141 LeuSerLysTrpAspProValValLeuLysAsnArgGlnAlaGluGlnLeu 157

RESULT 3
AAW21591
ID AAW21591 standard; peptide; 434 AA.
XX
AC AAW21591;
XX
XX 26-AUG-1997 (first entry)
DT
DE Antibiatic potentiating peptide #3.
XX
KW Potentiate; antibiatic; microbial infection; lipopolysaccharide;
KW permeability; outer bacterial membrane.
XX
OS Synthetic.
XX
PN WO9638163-A1.
XX
PD 05-DEC-1996.
XX
PF 29-MAY-1996; 96WO-EP02313.
XX
PR 31-MAY-1995; 95US-0456112.
XX
PA (BIOS-) BIOSYNTH SRL.
XX
PI Porro M, Varra M;
XX
DR WPI; 1997-034095/03.
XX
PT Potentiating activity of antibiatic with peptide contg. cationic
amino acid sequence - reduces dose of antibiatic required
XX
PS Disclosure; Page 7; 37pp; English.
XX
XX The sequences given in AAW21589-633 represent peptides which act to
potentiate the activity of an antibiatic when they are co-administered
with the antibiatic. Compositions containing these peptides are used
to treat or prevent microbial infections. These peptides bind to
lipopolysaccharide on the bacteria so may increase permeability of
the outer bacterial membrane to the antibiatic, allowing a reduction
in the dose of antibiatic required by 10-90% of the normal dose for
in vivo or in vitro application. Any toxic side effects are
correspondingly reduced.
XX
SQ Sequence 434 AA;
Alignment Scores:
Pred. No.: 9,93e-30 Length: 434
Score: 401.00 Matches: 79
Percent Similarity: 73.33% Conservative: 86
Best Local Similarity: 35.11% Mismatches: 56
Query Match: 18.74% Indels: 4
DB: 18 Gaps: 0
US-09-502-945-3 (1-1298) x AAW21591 (1-434)
QY 630 GAAAGGCTCTCTCCGAGCATGAGCTAGAGAGGCAAGATGCGACGAGAGAGCTT 689
Db 1 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 20
QY 690 CAGAGGCTCGGGCTCTGCAGTCTCTACTATGANGCCAGGCTCGAAGAGAGAGAAATC 749
Db 21 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
QY 750 NAAAGTTAAAGTATCAAAAGTCGTGAAGAAAGGAAAGCCCAAGAAAGCCCTAAAGAG 809
Db 41 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 60
QY 810 TTTGAGCAGCTCGGAGGTTTATCCAGCTGCCGACTAGAACCAACCAAGAGAGAA 869
Db 61 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
QY 870 AGAAGGAGGAGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 929

DR WPI; 2000-500381/45.
DR N-PSDB; AAC02146.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 6221; 71pp + CD-ROM; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 71 AA;

Alignment Scores:
Pred. No.: 8.72e-23 Length: 71
Score: 326.00 Matches: 68
Percent Similarity: 95.77% Conservative: 0
Best Local Similarity: 95.77% Mismatches: 3
Query Match: 15.23% Indels: 0
DB: 21 Gaps: 0

US-09-502-945-3 (1-1298) x AAG02140 (1-71)
QY 11 ATGACTGCGAACCGCTTGAGAGAGCTCTGCTTTGAGCCAGGAGGAGACTAGCG 70
Db 1 MetThrAlaAsnArgLeuAlaGluSerLeuLeuSerGlnGlnGluLeuLeuAla 20
QY 71 GATTGTCGCAAAAGACTACCTCTTGAGTGAGGTGAAGATGAGGGGACAATGATGGAGAG 130
Db 21 AspLeuProLysAspTyrLeuLeuSerGluSerGluAspGluGlyAspAsnAspGlyGlu 40
QY 131 AGAAGCATCNAAGCTCTGAGAGCAATCAGTCCCTTGATGGAAGATAGCGGAAA 190
Db 41 ArgLysHisGlnLysLeuLeuGluAlaIleSerSerLeuAspGlyLysAsnArgArgLys 60
QY 191 TTGGCTGANAAGTCTGAGGCTAGTCTGAAGGTG 223
Db 61 LeuAlaGluArgSerGluAlaSerLeuLysVal 71

RESULT 8
ABG26522
ID ABG26522 standard; Protein; 152 AA.
XX
AC ABG26522;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #26513.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Dmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS90709.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 56881; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC imaging of sites expressing (II). (I) and (II) are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 152 AA;

Alignment Scores:
Pred. No.: 1.24e-22 Length: 152
Score: 325.50 Matches: 66
Percent Similarity: 75.89% Conservative: 41
Best Local Similarity: 46.81% Mismatches: 24
Query Match: 15.21% Indels: 10
DB: 22 Gaps: 4

US-09-502-945-3 (1-1298) x ABG26522 (1-152)
QY 842 CGCAGTGAAGAGACGAG 901
Db 19 ArgArgArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluGly 38
QY 902 GAAGAAGAAGAAGAAG 961
Db 39 GluGlu-----GluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 54
QY 962 AGAAG 1021
Db 55 ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 74
QY 1022 GAGGAA---GAATAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1078
Db 75 GluGluGlyGluGlyGluGlyGluGlyGluGlyGluGlyGluGlyGluGlyGluGly 93
QY 1079 GAAGGAAGAAG 1138
Db 94 GluGlyArgArgArg-ArgArgArgArgArgArgArgArgArgArgArgArgArgArg 113
QY 1139 GGAACNAG 1198
Db 113 g-----ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 130


```
Db 130 AsnAspGlyAspArgGluGlyAspGluLysGluThrArgAspValGluAspGluGlyGlu 149
      :::::::::::::: :::::::::: :::: ::::
QY 155 -----GCAATCACTCCCTGTGAGAAAGAAATAGCGGAAATGGCTGAN 199
      :::::::::: ::::
Db 150 ThrArgArgSerSerPheAlaGluGlnThrGlyAsnGluArgThrGluMetArgThr 169
      :::::::::: ::::
QY 200 AGG----- 202
      ::::
Db 170 ArgHisGlyAspGluGlyTrpThrSerLysSerAsnArgPheAlaPheAlaCysPro 189
QY 203 -----TCTGAGCTAGTCTGAAGTGTGAGAGTTC 232
      :::::::::: ::::
Db 190 ArgPheSerLysSerAspValCysCysSerProGlnAlaArgLeuSerLeuProGluGln 209
QY 233 AATGTCAGTCTCAAGATCAGGAGAAAGCTGCTCTCCAGATCTGCTTGAGCCTGTT 292
      :::::::::: ::::
Db 210 SerLeuGlySerSerProSerProIleSerValThrAsnAspValTyrAlaLeuPhe 229
QY 293 AAAAATCTCATCTCTTTGGCCACTGTGAAAAGCAACTGAGTAGAGTCANATCAAGAAN 352
      :::::::::: ::::
Db 230 AspSerSerAlaSer----- 234
QY 353 ACAGTGGAGTTACTCTGAACAAGAAGAGATTGAACGGATCCACAGAGAAATAGCATTC 412
      :::::::::: ::::
Db 235 -----ProLeuHisAlaGlyGluLeuSerSerLeuProGlyAlaValSerAla 250
QY 413 ATAAAGCG-----ACAAGTCTCTCCAAATGGACCTGCTGCTGCTGAAGAACGGCA 466
      :::::::::: ::::
Db 251 SerGluArgLeuLeuThrAlaProAlaGluIleGlyProSerAlaSerSerAlaCysLeu 270
QY 467 GGCAGAGACAGCTGGTTTTTCCCTGGAGAAAGAGGAGCAGCCATTGCTCCCATTAACA 526
      :::::::::: ::::
Db 271 SerValSerCysGly-----ProGlyGluMet-SerProThrAlaAspThrArgHi 288
QY 527 TGRGCTCAGTGGCTGGAAGCAGCAACTCCCTGGAGCAGCAAAATTTCAACCTCTCCA 586
      :::::::::: ::::
Db 288 sAspAlaGluGluArgGluArgArgAlaGluGluGluLys-----G1 303
QY 587 TAGAACACAGCAGCCAGTACAGACCCCTTACTGACCCCTGTGAAAAGCCCTCTCTCCG 646
      :::::::::: ::::
Db 303 uArgGluArgGln-----GluGluGluGluArgGluArg 315
QY 647 AGCCATGAGCTAGAGAGGCAAGATGCGACGACGAGCTTCAGAGGCTCGGGCTCT 706
      :::::::::: ::::
Db 315 gArgValGluGluGluGluArgGluArgGluGluGluGluGluArgGluArg 335
QY 707 GCAGTCTACTATGANGCCAGGCTCGAAGAGAGAAATCNAAAGTTAAAGTATCA 766
      :::::::::: ::::
Db 335 gValGluGluGluLysAlaArgGlnArgGluGluGluGluGluGluGluGluArg 351
QY 767 CAAGTCTGTAAGAAAGAAAGGCAAGAAAGCCCTAAAGAGTTTCAGCAGCTCGCGAA 826
      :::::::::: ::::
Db 351 gArgArgValGluGluGluLysAlaArgGlnArgGluGluGluGluGluArgGluArg 371
QY 827 GGTTAATCCAGCTCCGCTAGAGAGCAAGCAAGAAAGAAAGAAAGAGAGAGAGAGA-- 884
      :::::::::: ::::
Db 371 gValGluGluGluLysAlaArgGlnArgGluGluGluGluGluGluGluGluArg 391
QY 885 -----AAGAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 940
      :::::::::: ::::
Db 391 lGluGluGluLysAlaArgGlnArgGluGluGluGluGluGluGluGluArgValG1 411
QY 941 AAGAGAGAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1000
      :::::::::: ::::
Db 411 uGluGluLysAlaArgGlnArgGluGluGluGluGluGluGluGluArgValGluG1 431
QY 1001 AAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1060
      :::::::::: ::::
Db 431 uGlu-----LysAlaArgGlnArgGluGluGluGluGluGluGluGluGluGlu 449
QY 1061 AAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120
      :::::::::: ::::
```

```
Db 449 uGluGluLysAlaArgGlnArgGluGluGluGluGluGluArgGluArgValGluGluG1 469
QY 1121 AAGAAAGAAAGAAATAAGGACNAGAAAGAAAGAAAGAAAGAAAG----- 1163
      :::::::::: ::::
Db 469 uLysGluArgGluArgGlnGluGluGluGluGluGluArgGluArgValGluGluLysG1 489
QY 1164 -AATAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAG 1222
      :::::::::: ::::
Db 489 uArgGluArgGlnGluGluGluGluGluGluGluGluArgGluArgValGluGluLysGluAr 509
QY 1223 AAAAAGAGAGAGAAAGTAGAAAGCGGAAGAAAGAAAGAAAGATATAAGAGAGAGAGAG 1282
      :::::::::: ::::
Db 509 gGluArgGlnGluGluGluLysArgGluArgGluGluGluGluGluGluLysAlaArgG1 529
QY 1283 AAGAAAGAGAGAGAGAA 1298
      :::::::::: ::::
Db 529 nArgGlnGluGluGlu 534
      :::::::::: ::::
RESULT 12
AAU25510
ID AAU25510 standard; Protein; 611 AA.
XX
AC AAU25510;
XX
DT 17-DEC-2001 (first entry)
XX
DE T. gondii immunogenic protein POC-23.
XX
KW Immunogenic protein; oocyst; faeces; enteric apicomplexa oocyst;
KW Cryptosporidium oocyst; Toxoplasma oocyst; Giardia cyst; vaccine;
KW oocyte shedding.
XX
OS Toxoplasma gondii.
XX
PN US2001014447-A1.
XX
PD 16-AUG-2001.
XX
PF 18-DEC-1998; 98US-0216393.
XX
PR 19-DEC-1997; 97US-0994825.
XX
PA (MILH/) MILHAUSEN M J.
XX
PI Milhausen MJ;
XX
DR WPI: 2001-529100/58.
DR N-PSDB; AAS42565.
XX
PT Detecting parasite oocysts or cysts in faeces, comprises eluting DNA
PT from sample into aqueous solution by heating, amplifying DNA with
PT primers specific for oocysts or cysts being detected, and detecting
PT amplification product.
XX
PS Example 2; Page 91-93; 188pp; English.
XX
CC The invention relates to detection of parasite oocysts or cysts in
CC a faeces sample comprising contacting the sample with a solid support,
CC drying and then washing the sample with an aqueous wash solution, adding
CC an aqueous elution solution and eluting DNA from the sample by heating
CC and amplifying by PCR oocyst/cyst-specific DNA and detecting the
CC amplification products. The method is useful for detecting parasite
CC oocysts e.g., enteric apicomplexa oocysts such as Cryptosporidium oocysts
CC or Toxoplasma oocysts, or for detecting parasite cysts e.g. Giardia
CC cysts. The method is also useful for developing vaccines to prevent
CC oocyte shedding in cats. The present sequence represents an
CC immunogenic protein from Toxoplasma gondii.
SQ Sequence 611 AA;
Alignment Scores: 3.05e-21 Length: 611
Pred. No.: 313.00 Matches: 100
Score:
```


D	b		411 uGluGluLysAlaArgGlnArgGluGluGluGluGluArgGluGlyArgValGluGluG	431
Q	y	1001	RAGAAGAACTAAGAAGAAGAGAGAGGAATAGAAGAGGAAGAAAAGAAAAAGTNA	1060
D	b	431	uGU-----LysAlaArgGlnArgGluGluGluGluGluArgGlyArgValGL	449
Q	y	1061	AGAAGAAAAGGAAGAAGAAAGAAAGAGAGGAACCTNAGAAAGAAAGAGGAGGA	1120
D	b	449	uGluGluLysAlaArgGlnArgGluGluGluGluGluArgArgValGluGluGluG	469
Q	y	1121	AAGAAGAAGAAAGTAAGAACAACNAGAAAGAGGAGGAAGAAAG-----	1163
D	b	469	uLySgluArgGlnGluGluGluGluGluGluArgArgValGluGluGluLysGL	489
Q	y	1164	-AATAAGAAGAGGAGAGAAAAGAAAGAAAGAGAGGAAGAGAGAAAAGGAAG	1222
D	b	489	uArgGluArgGlnGluGluGluGluGluArgGluArgValGluGluGluLysGluAr	509
Q	y	1223	AAAAAAGGAAGAAGACTAGAAGCGGAAGAAAGAAAGAAAGATATAAGAGGAAG	1282
D	b	509	gGluArgGlnGluGluLysArgGluArgArgValGluGluGluLysAlaArgGL	529
Q	y	1283	AAGAAGAAGAAAAA 1298	
D	b	529	nArgGlnGluGlu 534	
RESULT 13				
ABB58657				
ID	ABB58657 standard; Protein; 2274 AA.			
XX	AC	ABB58657;		
DT	26-MAR-2002	(first entry)		
DE	Drosophila melanogaster polypeptide SEQ ID NO 2763.			
KW	Drosophila; developmental biology; cell signalling; insecticide;			
KW	pharmaceutical.			
OS	Drosophila melanogaster.			
PN	WO200171042-A2.			
PD	27-SEP-2001.			
Pf	23-MAR-2001; 2001WO-US09231.			
PR	23-MAR-2000; 2000US-191637P.			
PR	11-JUL-2000; 2000US-061415O.			
PA	(PEKE) PE CORP NY.			
PI	Venter JC, Adams M, Li PWD, Myers EW;			
DR	WPI; 2001-65860/75.			
DR	N-PSDB; ABL02760.			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more			
PT	genes from Drosophila and for elucidating cell signalling and cell-cell			
PT	interactions .			
PS	Disclosure; SEQ ID NO 2763; 2lpp + Sequence Listing; English.			
CC	The invention relates to an isolated nucleic acid detection reagent			
CC	capable of detecting 1000 or more genes from Drosophila. The invention			
CC	useful in developmental biology and in elucidating cell signalling and			
CC	cell-cell interactions in higher eukaryotes for the development of			
CC	insecticides, therapeutics and pharmaceutical drugs. The invention			
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA			
CC	sequences (ABL01840-ABLL16175) and the encoded proteins			
CC	(ABB57737-ABB72072).			
CC	The sequence data for this patent did not form part of the printed			
CC	specification, but was obtained in electronic format directly from WIPO			

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2274 AA;

Alignment Scores:

Pred. No.: 5,77e-21 Length: 2274
Score: 312.00 Matches: 128
Percent Similarity: 47.31% Conservative: 83
Best Local Similarity: 28.70% Mismatches: 136
Query Match: 14.58% Indels: 99
DB: 22 Gaps: 16

US-09-502-945-3 (1-1298) x ABB58657 (1-2274)

```
QY 101 AGTGAACATGAGGGGACATGATGGAGAGAGAAAGCATCNAAGCTTCTGGAGCAATC 160
    |||||
Db 1203 SerGluAspGluAlaThr-----ThrThrIleThrLeuAlaLys 1215
QY 161 AGTTCCTTGTATGGAAGAATAGCGGAAATTGGCTGANAGGCTGAGGCTAGTCTGAAG 220
    |||||
Db 1216 SerAlaLysLysAlaLysAsnAsnLysGln-----LysThrSerGlyLys 1230
QY 221 GTGTCAGATTCATGTCAGTCTCGAAGGATCAGGAGAAAGCTGGCTCTGCAGATCTG 280
    |||||
Db 1231 ValSerGluLysLysProGluGlnThrAsnLeuAlaGluGluLysVal----- 1246
QY 281 CTTCAGCTGTAAACTTCATCTCTTGGCCACTCTGAAAGCAACTCAGTAGAGTC 340
    |||||
Db 1247 -----AlaValGluLysGluGluAsnGluGlu 1255
QY 341 ANATCAAGANACAGTGGAGTTACCTCTGAACAAAGAA-----GAGATTGAACGG 391
    |||||
Db 1256 ArgAlaValProMetValLysLysThrIleAsnLysGluGlnAspSerAspAlaGluSer 1275
QY 392 ATCCACAGAGAAATGATTCATAAAGCGACAAAGTCTCTCCAAATGGGACCTGTCGT 451
    |||||
Db 1276 AspHisAlaAsp-SerLeuLeuAlaAsnLysSerSerIleAla-----AlaValMe 1292
QY 452 CCTCAAGAACCGGCAGCAGCAGCAGCTGTTTTCCTCGGAGAAAGAGGAGCCAGCAT 511
    |||||
Db 1292 tValSerSerAlaSerAlaGlnGlyLeuSerLeuHisValGlu----- 1306
QY 512 TGCTCCCATTTGAATGCTGCTAGTGGCTGGAAGCGCAAGAACTCCCTGGAGCAGGAAT 571
    |||||
Db 1307 -----MetSerAlaAlaAspAlaGluGlnGlyGluAspGluGluI 1320
QY 572 TTTCAACCTC-----CTCCATAGAACACAGCAGCC 601
    |||||
Db 1320 eGluGlyLeuAspGluGluProLysThrMetSerLysAspAsnLysLysGlnL 1340
QY 602 AGTGACAGACCTTTACTAGCCCTGTG-----GAAAAGGCTCTCTCCG 646
    |||||
Db 1340 sProGlyAspAlaValAlaThrMetThrIleAspLysGluLysGluLysAlaLysGluL 1360
QY 647 AGCCATGACCTTAGAAGAGCAAGATCGACGACGAGCTTCAG----- 692
    |||||
Db 1360 sGluLeuLysLysGluLysGluLysGluLysGluLysGluLysGluLysGluLysG 1379
QY 693 -----AGGGCTCGGGCTCT 706
    |||||
Db 1379 uGluLysLeuLysLeuLysGluLysGluLysGluLysLeuArgMetGluLysGluLysLe 1399
QY 707 GCAGTCTTACTATGANGCAGGCTCGAAGAGAGAGAAATCNAAGTTAAAGATCA 766
    |||||
Db 1399 uLysGluGluLysLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 1416
QY 767 CAAGTCTGTGAAGAAAGAAAGCAAGAGCCCTAAAGAGGTTTGAGCAGCTCGCGAA 826
    |||||
Db 1416 uLysIleLysGluLysGlnArgGluGluLys--LeuLysGlu--GluLysLeuLysG 1434
QY 827 GGTAAATCCAGTCCGCTAGAGAACGAGAAAGAGGAAAGA-----AGGAG 877
    |||||
Db 1434 uLysGluArgGluGluArgMetLysGluLysGluArgGluGluLysAlaLysGluLysG 1454
```

```
QY 878 GAGGACAAAGAAAGAAAGGAGAAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 937
    |||||
Db 1454 nArgGluLysLysLeuArgGluGluLysLysLysLysLysLysLysLysLysLysLysLys 1474
QY 938 AGGAGAAAGGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 985
    |||||
Db 1474 ysGluLysLeuArgGluGluLysLysLysLysLysLysLysLysLysLysLysLysLys 1494
QY 986 AAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1045
    |||||
Db 1494 luArgGluGluLysMetArgGlu--LysGluArgGluGluLysLysLysLysLysLysLys 1513
QY 1046 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1105
    |||||
Db 1513 gValGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1533
QY 1106 GAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1165
    |||||
Db 1533 sLeuLysGluLysGluGluLeuLeuLysLysLysLysLysLysLysLysLysLysLysLys 1553
QY 1166 TRAG---AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1222
    |||||
Db 1553 uLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1573
QY 1223 AAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1282
    |||||
Db 1573 uLysArgGluThrGluGluArgGlnArgGluLysGluLysGluLysLysLysLysLysLys 1593
QY 1283 AAGAAAGA 1290
    |||||
Db 1593 ysGluArg 1595
RESULT 14
ABB62436
ID ABB62436 standard; Protein; 771 AA.
XX
AC ABB62436;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 14100.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WIPI; 2001-656860/75.
XX
DR N-PSDB; ABL06539.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 14100; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
```


CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (AB157737-AB172072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 771 AA;

Alignment Scores:

Pred. No.:	7,11e-21	Length:	771
Score:	309.50	Matches:	123
Percent Similarity:	45.15%	Conservative:	77
Best Local Similarity:	27.77%	Mismatches:	180
Query Match:	14.46%	Indels:	63
DB:	22	Gaps:	15

US-09-502-945-3 (1-1298) x ABB62436 (1-771)

QY	101	AGTGAAGATGAGGGGCAACATGATGGAGAGAAAGCATCNAAGCTTCTGGAACATC	160
Db	2	SerAspAspGlu---GluHisTyrGluProLysAlaHisLysLysLeuLeuGlnAlaIle	20
QY	161	AGTTCCTTGATGGAAGCAATAGCGGAAATGGCTGANAGGCTGAGGCTAGTCTGAAG	220
Db	21	GlySerLeuGlyLysValGlnHisIleGlnLysSerThrArgAspGluArgGlnSerLeu	40
QY	221	GTGTCAGAGTTC-----AATGTCAGTCTCTGAAGGATCAGGAGAA	259
Db	41	GlnAspGluPheGlnLeuValLysGlyValSerSerAlaGluThrAspHisAlaProArg	60
QY	260	AGCTGTCCTTGAGATCTGCTTGAGCCTGTTAAACTTCATCTCTTTGGCCACTGTG	319
Db	61	AlaValGlyLeuAsnAspLeuValAspIleLeuArgThrSerThrLysHisSerGlnThr	80
QY	320	AAAAGCAACACGACGACGATC---ANATCAAGAAACAGTGGAGTTACCTCTGAACAA	376
Db	81	GlyLysLysLeuLysAsnIleHisGlySerLysValLysValLeuGlnLysProLeuGluLys	100
QY	377	GAAGAGATTGAACGATCCACAGAGATAGCATTCATCAATAAAGCAGCAGATC-----	428
Db	101	ProAlaAlaAspArg-----IleLysArgThr-IleGlyTyrGln	113
QY	429	-----CTCTCCAAATGAGCCCTGCTGCTGCTGAAGACCGCAGGAGAGA	472
Db	113	uGlyValThrLysLysLysLeuGlyArgTyrAspAlaValAlaGlnGlnArgSerAlaGln	133
QY	473	GCAGCTGCTTTTCCCTGGGAGAAGAGGAGCCAGCCATCTCTCCATTGAACATGTGCT	532
Db	133	uthrGlnIlePheProLeuProSerGluThrValTyrValAsnThrAlaAlaAsnAlaAr	153
QY	533	CAGT---GGCTGGGAAGCAAGCACTCCCTGGAGCAGGAATTT-----TTCAA	577
Db	153	gProLeuAsnThrArgValLysSerAsnLeuAlaGlnGluLeuAlaAsnAsnArgLys	173
QY	578	CCTCTCCATGAACAGCAGCAGCTGACAGCCCTTTA-----CTGACCCC	625
Db	173	LeuArgGluLeuArgLysAlaGlnIleGlyAspThrThrAspGluLysGluLeuAlaLys	193
QY	626	TGTGGAAGGCCCTCTCTC---CGAGCCATGAGCTGAGAGCAAGATGCGACGAGC	682
Db	193	sGlnGluArgLeuLeuGlnLysLysLeuThrArgAspGluLeuPheAlaArgArgLys	213
QY	683	AGAGCTTCAGAGGCTCGGGCTCTGCTGCTCTACTATGANGCAAGCTCGAAGAGAGAA	742
Db	213	sGluLeuAlaTyrLeuLysMetArgGluSerGlnLysSerAlaLysAlaArgMetGlnAs	233
QY	743	GAATAATCAAGTTAAAGTATCAAAAGTGTGAAGAAAGGAAGGCCAAGAACGCCCT	802
Db	233	nLysIleLysSerLysLysPheHisLysLeuGlnLysArgGlnLysMetLeuGlnGlnMe	253

QY	803	AAAAGAGTTTCAGCAGCTGCGGAAGGTTAATCCAGCTGCC-----GC	844
Db	253	tLysGluPheGluLeuLeuGlnLysThrAspProGluAlaAlaLeuGluLysLeuAsnAl	273
QY	845	ACTAGAAGAACGAAGAAAGAGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	904
Db	273	aLeuGluLysSerArgValGlnGluArgAlaSerLeuArgHisLysAsnThrGlyThrTr	293
QY	905	GAAGAAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	964
Db	293	pAlaLysAsnLeuGlnIleArgAlaLysTyrAspLysAspValArgLysAspLeuAlaGln	313
QY	965	AAGAAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1024
Db	313	uGlnLeuAlaValSerArg-----GluLeuThrGlnLysLysGlnValSe	328
QY	1025	GAAGAATAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1078
Db	328	rAspSerGluAspGluThrThrLysLysArgValLeuProGluGluAspAsnAspTy	348
QY	1079	GAAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1138
Db	348	rAspProPheAsnProThrThrLysArgLysAlaThrGluGlnThrGluAsnGlyAsnGln	368
QY	1139	GGACACNAGAA-----AGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1174
Db	368	yGluSerGluGlySerAsnTrpArgGlnTyrTrpThrLysArgAsnLysAsnGluLysLe	388
QY	1175	GAAGAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1216
Db	388	uLeuGlu-GluHisArgArgAspLeuGluGlyValGlnAspAsnAspGluGlnGluLul	408
QY	1217	AGGAAGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1276
Db	408	yGluThrIleArgLysProLeuIleGluValLysProAlaLysLysLysLysLysLys	428
QY	1277	AAGAA 1281	
Db	428	erLys 429	
RESULT 15			
ID	ABG26543		
XX	ABG26543 standard; Protein; 405 AA.		
AC	ABG26543;		
XX			
DT	18-FEB-2002 (first entry)		
XX			
DE	Novel human diagnostic protein #26534.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder.		
XX			
OS	Homo sapiens.		
XX			
PN	W0200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001WO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmanac RT, Liu C, Tang YT;		
XX			
DR	WPI; 2001-639362/73.		
DR	N-PSDB; AAS90730.		
XX			
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		

PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX
2255

PS Claim 20; SEQ ID No 56902; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (III). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AEG00010-AEG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 405 AA;

Alignment Scores:

Pred. No.:	6.47e-21	Length:	405
Score:	309.00	Matches:	83
Percent Similarity:	59.75%	Conservative:	61
Best Local Similarity:	34.44%	Mismatches:	56
Query Match:	12.44%	Indels:	41
DB:	22	Gaps:	7

US-09-502-945-3 (1-1298) x ABG26543 (1-405)

Qy	633	AAGCC	TCTCTCCGAGCCATGACGCTAGAAAGAGAGCAAGATCGGACGACGACAGACTTCAG	692
Db	148	GLuAlaProLeuAlaCysMetAlaAlaGluGluThrCysIleGluGlyAlaSerThrVal	167	
Qy	693	AGGCTCGG-----GCCTCGAGCTCTACTCATGANGCCAAAGGCT	731	
Db	168	AspAlaArgGluProPheLeuGluValLeuLeuArgTyrProPheProGlnGly	187	
Qy	732	-----CGAAGAGAGAGAAATCNAAAGTTATAAGTAGT	764	
Db	188	ProPheGluGlyProGlnArgAlaPheLysThrLysAsnLysThrGlnGlyLeuGlnTyr	207	
Qy	765	CACAAAGTCGTGANAAGGAAGGCCAACAAAGCC-----CTAAAAAGAGTTTGG	815	
Db	208	LeuArgMetLeuSerProArgGlySerLysLysSerGlnCysLysValThrGluPheLeu	227	
Qy	816	CAGCTG-----CGAAGGTTAATCCAGCTGCCGCACTA	848	
Db	228	GluLeuValGlnAspGlnGlyLysAspAsnHisArgAlaIleSerProGlnArgGlyArg	247	
Qy	849	GAAGACGACGAAGAAAGGAGAAAGGAGGAGGAGAAAGAAGAACCAAGGAGAGAAAG	908	
Db	248	ArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg	267	
Qy	909	AAAGAAGAAAGGAGAGAGAGAAAGAAGAGGAGAGAGAGAAAGGAAGAGAGAGAGA	968	
Db	267	gArgArgArg-----ArgArgLysLysLysLysLysLysLysLysLysLysLysLys	283	
Qy	969	AAAGGAGAAAGAAAGAAAGGAGAGAGAAAGAAAGAACTAAGAAAGAGGAGAGAGAG	1021	
Db	283	sluysl	303	
Qy	1029	AATAAGAGGAGAAAGAAAGAAATAAGTTNAAAGAG---AAGAAAGAGAGAGAGAA	1081	

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:34:32 ; Search time 10.0556 Seconds
(without alignments)
7595.986 Million cell updates/sec

Title: US-09-502-945-3
Perfect score: 2140
Sequence: 1 ggcgtggaatgactgcga.....gaagaagaagaagggaaaaa 1298

Scoring table: BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p_model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09502945/runat_14032003_101059_19145/app_query.fasta_1.10979
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_80 -runat_14032003_101059_19145 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/6C.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	401	18.7	434	1	US-08-097-830B-3
2	401	18.7	434	2	US-08-456-112B-3
3	298	13.9	1898	1	US-08-056-200-94
4	298	13.9	1898	2	US-08-800-644-94
5	265	12.4	100	2	US-08-460-890A-62
6	265	12.4	100	3	US-08-167-641C-62
7	265	12.4	100	4	US-08-460-971A-62
8	265	12.4	100	4	US-08-462-040-62
9	261	12.2	1507	3	US-08-929-329-5
10	234.5	11.0	432	2	US-08-933-750C-47
11	234.5	11.0	432	4	US-09-234-613-47
12	211	9.9	558	1	US-08-285-440-6

13	211	9.9	558	1	US-08-630-349-6
14	209	9.8	257	4	US-09-461-697-188
15	209	9.8	272	4	US-09-461-697-186
16	205	9.6	231	4	US-09-461-697-194
17	205	9.6	232	4	US-09-461-697-192
18	205	9.6	238	4	US-09-461-697-190
19	201	9.4	532	1	US-08-285-440-5
20	201	9.4	532	1	US-08-630-349-5
21	200.5	9.4	300	4	US-09-395-689-1
22	200.5	9.4	765	2	US-08-663-112-2
23	198.5	9.3	1588	5	PCT-US93-07261-11
24	198.5	9.3	1663	5	PCT-US93-07261-16
25	189	8.8	683	6	5210183-3
26	188.5	8.8	290	2	US-08-903-801-1
27	188.5	8.8	290	4	US-09-295-055-1
28	188.5	8.8	614	5	PCT-US95-03236-21
29	185.5	8.7	288	3	US-08-961-858-6
30	185.5	8.7	288	3	US-09-089-593-6
31	185.5	8.7	288	3	US-08-950-925-4
32	180.5	8.4	1360	4	US-09-393-569-2
33	179.5	8.4	118	4	US-09-134-001C-2856
34	179.5	8.4	316	4	US-08-098-327E-31
35	179.5	8.4	316	4	US-08-462-625-31
36	178	8.3	631	4	US-08-847-065-25
37	177	8.3	234	2	US-08-903-801-3
38	177	8.3	234	4	US-09-295-055-3
39	177	8.3	564	4	US-09-308-022-6
40	176.5	8.2	816	2	US-08-533-306A-6
41	176.5	8.2	816	2	US-08-742-923A-6
42	176.5	8.2	885	2	US-08-533-306A-4
43	176.5	8.2	885	2	US-08-742-923A-4
44	175	8.2	514	2	US-08-960-022-14
45	174	8.1	564	2	US-08-216-894-2

ALIGNMENTS

RESULT 1
US-08-097-830B-3
; Sequence 3, Application US/08097830E
; Patent No. 5652211
; GENERAL INFORMATION:
; APPLICANT: Porro, Massimo
; TITLE OF INVENTION: Peptides For Neutralizing The
; TITLE OF INVENTION: Toxicity of Lipid A
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hedman, Gibson & Costigan, P.C.
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/097,830E
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Costigan, James V.
; REGISTRATION NUMBER: 25,669
; REFERENCE/DOCKET NUMBER: 576-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 302-8989
; TELEFAX: (212) 302-8998
; INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 434 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-097-830E-3

Alignment Scores:
Pred. No.: 8.34e-32 Length: 434
Score: 401.00 Matches: 79
Percent Similarity: 73.33% Conservative: 86
Best Local Similarity: 35.11% Mismatches: 56
Query Match: 18.74% Indels: 4
DB: 0 Gaps: 0

US-09-502-945-3 (1-1298) x US-08-097-830E-3 (1-434)

```
QY 630 GAAAGGCGCTCTCCGAGCCATGAGCTAGAGAGGCAAGATCGCAGCAGCAGCTT 689
Db 1 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 20
QY 690 CAGAGGCGTGGGCTCTGCGAGTCTACTATGANGCCAGGCTCGAAGAGAGAAATC 749
Db 21 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
QY 750 NAAAGTTAAAGATCATCAAGTCTGNAGAGAGAGAGGCGCAAGAGCCCTAAAGAG 809
Db 41 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 60
QY 810 TTTGAGCAGCTGCGGAGGTTAATCCAGCTCGCCAGCTAGAGAGAGAGAGAGAA 869
Db 61 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
QY 870 AGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 929
Db 81 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100
QY 930 AGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 989
Db 100 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 120
QY 990 GAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
Db 120 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 139
QY 1050 AAAAAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1109
Db 140 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 159
QY 1110 AAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1169
Db 160 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 179
QY 1170 AAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1229
Db 180 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 199
QY 1230 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1289
Db 199 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 219
QY 1290 AAGGAAA 1296
Db 219 sLysLys 221
```

RESULT 2

US-08-456-112B-3
Sequence 3, Application US/08456112B
Patent No. 5834430
GENERAL INFORMATION:
APPLICANT: Porro, Massimo
TITLE OF INVENTION: POTENTIATION OF ANTIBIOTICS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:

ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: LEADING EDGE 486
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,112B
FILING DATE: May 31, 1995
CLASSIFICATION: 514
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 434 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-08-456-112B-3

Alignment Scores:
Pred. No.: 8.34e-32 Length: 434
Score: 401.00 Matches: 79
Percent Similarity: 73.33% Conservative: 86
Best Local Similarity: 35.11% Mismatches: 56
Query Match: 18.74% Indels: 4
DB: 0 Gaps: 0

US-09-502-945-3 (1-1298) x US-08-456-112B-3 (1-434)

```
QY 630 GAAAGGCGCTCTCCGAGCCATGAGCTAGAGAGGCAAGATCGCAGCAGCAGCTT 689
Db 1 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 20
QY 690 CAGAGGCGTGGGCTCTGCGAGTCTACTATGANGCCAGGCTCGAAGAGAGAAATC 749
Db 21 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40
QY 750 NAAAGTTAAAGATCATCAAGTCTGNAGAGAGAGAGGCGCAAGAGCCCTAAAGAG 809
Db 41 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 60
QY 810 TTTGAGCAGCTGCGGAGGTTAATCCAGCTCGCCAGCTAGAGAGAGAGAGAGAA 869
Db 61 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
QY 870 AGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 929
Db 81 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100
QY 930 AGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 989
Db 100 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 120
QY 990 GAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
Db 120 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 139
QY 1050 AAAAAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1109
Db 140 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 159
```



```

QY 1238  AAGTAAAGCGGAGAAAGAAAGAAAGTATAGAGAGGAGAGAAAGAAAGG 1299
      || ||| |||:|||||:|||||:|||||:|||||:|||||:|||||
Db 456  ySHisGluGlnGluArgGluGlnArgLeuLysArgGluGlnGluArgArg 474

RESULT 5
US-08-460-890A-62
; Sequence 62, Application US/08460890A
; Patent No. 5994109
; GENERAL INFORMATION:
; APPLICANT: Woo, Savio L.C.
; APPLICANT: Smith, Louis C.
; APPLICANT: Cristiano, Richard J.
; APPLICANT: Gottchalk, Stephen
; TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND
; TITLE OF INVENTION: METHODS OF USE
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,890A
; FILING DATE: June 5, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167,641
; FILING DATE: December 14, 1993
; APPLICATION NUMBER: 07/855,389
; FILING DATE: March 20, 1992
; APPLICATION NUMBER: PCT/US93/02725
; FILING DATE: March 19, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 212/066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: "Lys" in positions 3 to 100 may be present
; OTHER INFORMATION: or absent.
US-08-460-890A-62

Alignment Scores:
Pred. No.: 1,76e-18 Length: 100
Score: 265.00 Matches: 48
Percent Similarity: 89.32% Conservative: 44
Best Local Similarity: 46.60% Mismatches: 8
Query Match: 12.38% Indels: 3
DB: 2 Gaps: 0

US-09-502-945-3 (1-1298) x US-08-460-890A-62 (1-100)

QY 934  AAGAAGAGAGAGGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9
      |||||:|||||:|||||:|||||:|||||:|||||:|||||

```



```

Db 183 AlaAlaGlnGluLysGlnMetGluValCys----- 192
QY 545 GGCAAGAACTCCCTGGAGCAGGAAATTTCAACCTCTCCATAGAACAGCAGCCAGT 604
Db 192 ----- 192
QY 605 GACAGACCCCTTACTGACCCCTGTGGAAGGCCCTCTCCGAGCCATGAGCCTAGAAGA 664
Db 193 -----GluValCysGlyAlaHeLeuIleValGly----- 202
QY 665 GGCAAGAACTGGAGCAGCAGCAGCTTCAGAGCGCTCGGCTCTGCAGCTCTACTATGANGC 724
Db 203 -----AspAlaGlnSerArgValAspHis-----Leu 212
QY 725 CAAGGCTCGAGAGAGAGAAATCNAAGTTAAA-----GTATCACAAGTCGGA 778
Db 213 MetGlyLysGlnHisMetGlyTyrAlaLysIleLysAlaThrValGluLeuLysGlu 232
QY 779 GAA---AGGAAAGCCCAAGAGCCCTAAAAGAGTTTGAGCAGCTCGGAGGTTTATCC 835
Db 233 LysLeuArgLysArgThrGluGluProAsp----- 242
QY 836 AGTGGCCGCTAGAGAAGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 895
Db 243 -----ArgAspGluArgLeuLysLysGluLysGlnGluArgGluGluArg 257
QY 896 CAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 955
Db 258 GluLysGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGlu 277
QY 956 AGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1015
Db 278 GluArgGluLysGluArgAlaArgAspArgGluArgLysArgSer-ArgSerArgSe 297
QY 1016 GAAG-----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1057
Db 297 rArgHisSerSerArgThrSerAspArgArgCysSerArgSerArgAspHisLysArgSe 317
QY 1058 NAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
Db 317 rArgSerArgGluArgArgThrArgSerArgAspArgArgSerArgSerHisAs 337
QY 1118 G-----GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1171
Db 337 pArgSerGluArgLysHisArgSerArgSerArgSerArgSerArgSerArgSerLys 357
QY 1172 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1231
Db 357 pArgLysSerTyrLysHisArgSerLysSerArgAspArgGluGlnAspArgLysSer 377
QY 1232 AGAAGAACTAGAACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1285
Db 377 sGluLysGluLysArgGlySerAspAspLysLysSerValLysSerGlySerArgG1 397
QY 1286 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298
Db 397 uLysGlnSerGlu 401

RESULT 11
US-09-234-613-47
; Sequence 47, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMNOT04
; CLONE: 2926777
; US-09-234-613-47

Alignment Scores:
Pred. No.: 3,39e-15 Length: 432
Score: 234.50 Matches: 91
Percent Similarity: 41.23% Conservative: 76
Best Local Similarity: 22.47% Mismatches: 128
Query Match: 10.96% Indels: 110
DB: 4 Gaps: 15

US-09-502-945-3 (1-1298) x US-09-234-613-47 (1-432)
QY 149 CTGGAAGCAATCAGTTCCTTCCTGATGGAAGAAATAGCGGAAATTGGCTGANNAGTCTGAG 208
Db 86 LeuArgTyrLeuGlnSerLeuLeuAlaGluValGluArgGile---ArgArgGlyHis 104
QY 209 GCTAGTCTGAAGTGTCAGAGTTCAATGTCAGTTCGAGGATCAGGA----- 256
Db 105 AlaArgLeuAlaLeuSerGlnAsnGlnGlnSerSerGlyAlaAlaGlyProThrGlyLys 124
QY 257 -----GAAAGCTG---GTCCTTGCAGATCTGCTTGAGCTGTTAAACCTTCATCTTCT 307
Db 125 AsnGluGluLysIleGlnValLeuThrAspLysIleAsp----- 137
QY 308 TTGGCCACTGTGAAAGCAACTGAGTACAGTCANATCAAGAAANACAGTGGAGTTACCT 367
Db 138 -----ValLeuLeuGlnGlnIleGluLeuLysSerGlyLysValGluAla 155
QY 368 CTGAACAAAGAGAG---ATTGAACGATCCACAGAGATAGCATTAATAAACCCACA 424
Db 156 GlnGlyMetMetLysLeuValGluGlnLysGluGluArgGluLeuArgSerThr 175
QY 425 AGTCCTCTCCAAATGGGACCCTGCTGCTCTGAAGAACCGCAGGACAGCAGCTGTTT 484
Db 176 Thr-----SerThrIleGluSerPhe 182
QY 485 TCCCTCGGAGAAAGAGAGGAGCCCATTTGCTCCCATTTGAATGTGCTCAGTGGTGAA 544

```


Db 183 AlaAlaGlnGluLysGlnMetcluValCys----- 192
QY 545 GCACAAGACTCCCTCGGAGCAGGAATAATTTCAACCTCTCCATAGAACAAGCAGCCAGT 604
Db 192 ----- 192
QY 605 GACACACCTTTACTGACCCCTGTGAAAGGCGCTCTCCGAGCCATGAGCCTAGAGA 664
Db 193 ----- 202
QY 665 GCACAAGATGCGACGAGCAGCAGCTTCAGAAGGCTCGGCTCTGCAGCTCTACTATGANGC 724
Db 203 -----AspAlaGlnSerArgValAspHis-----Leu 212
QY 725 CAAGCTCGAAGAGAGAGAAATCNAAGTTAAA-----GTATCACAAGTCGTAA 778
Db 213 MetGlyLysGlnHisMetGlyTyrAlaLysIleLysAlaThrValGluGluLeuLysGlu 232
QY 779 GAA---AGGAAAGGCCAAGAAAGCCCTAAAGAGTTTGACGAGCTGCGGAGGTTAATCC 835
Db 233 LysLeuArgLysArgThrGluLysProasp----- 242
QY 836 AGCTGCCGCTAGAGAGACGAG 895
Db 243 -----ArgAspGluArgLeuLysLysGluLysGlnGluArgGluGluArg 257
QY 896 CAAGGAG 955
Db 258 GluLysGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArg 277
QY 956 AGAAGGAG 1015
Db 278 GluArgGluLysGluArgAlaArgAspArgGluArgLysArgSerArgSerArgSe 297
QY 1016 GAAG-----GAG 1057
Db 297 rArgHisSerArgThrSerAspArgArgCysSerArgSerArgSerArgHisLysArgSe 317
QY 1058 NAAAG 1117
Db 317 rArgSerArgGluArgArgArgThrArgSerArgSerArgSerArgSerArgSerHis 337
QY 1118 G-----GAAAG 1171
Db 337 pArgSerGluArgLysHisArgSerArgSerArgSerArgSerArgSerArgSerArg 357
QY 1172 GAG 1231
Db 357 pArgLysSerTyrLysHisArgSerLysSerArgAspArgGluGlnAspArgLysSerLy 377
QY 1232 AGAAG 1285
Db 377 sGluLysGluLysArgGlySerAspAspLysLysSerSerValLysSerGlySerArgG 397
QY 1286 AAAAGAGAGAGAGAG 1298
Db 397 uLysGlnSerGlu 401

RESULT 12
US-08-285-440-6
; Sequence 6, Application US/08285440
; Patent No. 5532337
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.

Alignment Scores:
Pred. No.: 8.25e-13
Score: 211.00
Percent Similarity: 40.17%
Best Local Similarity: 25.14%
Query Match: 9.86%

Length: 558
Matches: 87
Conservative: 52
Mismatches: 88
Indels: 119

ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: DisplayWrite
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/285,440
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/858,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-6

[illegible]

```

Db      372 uileGluArgAarg 377          I   III   IIIIII
RESULT 13
US-08-630-349-6
; Sequence 6, Application US/08630349
; Patent No. 5739008
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,349
; FILING DATE: April 10, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/285,440
; FILING DATE: August 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 558 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-630-349-6

Alignment Scores:
Pred. No.:      8.25e-13              Length:      558
Score:          211.00                Matches:     87
Percent Similarity: 40.17%           Conservative: 52
Best Local Similarity: 25.14%        Mismatches:  88
Query Match:     9.86%               Indels:     119
DB:                                     Gaps:       12

US-09-502-945-3 (1-1298) x US-08-630-349-6 (1-558)
Qy    579 CTCCTCCATAGAACAGACGCCAGTGCACAGACCTTTACTGACCCCTGTGGAAAGGCC 638
      III   III   ::::~             :::~             IIIII
Db    44 LeuArgGlnLysGlnGluGluSerLeuGlycInValThrAspGlnValGluValasn 63
Qy    639 TCTCTCGAGCATTAGACCTTAGAAGAGCAAGAAGATGCGACGACAGACTTCAG----- 692
      ::~   ::::~             |||||            |||
Db    64 AlaGlnAsnSerValProAspGluGluAlaLysThrThrThrAsnThrGlnValGlu 83
Qy    693 -----AGGGCTCGGCTCTCAGCTCTACTATCATGCCCAAGGCTCCAAGAGAGAAG 803
      |||   |||   :::~             :::~             :::~
Db    84 GlyaspAspGluAlaAlaPheLeuGluGluAlaArgArgGluGluArgGlnLys 103
Qy    744 AAAATCNAAAGCTTAAAAGTATCACAAAGTCGTGAAGAAAGGAAGGCCAAGAAGCCCTA 803
      ::::~             |||   :::~             :::~
Db    104 ArgLeuGln-----GluAlaLeuGluArgGln 112
Qy    804 AAAGAGTTTGAG-----CAGCTGCGCAAGGTT 830
      |||||            :::~

```



```
Db 113 LysGluPheAspProThrIleThrAspAlaSerLeuSerLeuProSerArgArgMetGln 132
QY 831 AATCCAGTCCCGCACTAGAAAGACGAAAGAAAGAGGAGGAGGAGAAAG--- 887
Db 133 AsnAspThrAlaGluAsnGluThrGluLysGluGluLysSerArgGlnGlu 152
QY 888 -----AAGAAAGAAAGGAGAG 905
Db 153 ArgTyrGluIleGluGluThrValThrLysSerTyrGlnLysAsnAspTrpArg 172
QY 906 AAGAAGAAAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA--- 959
Db 173 AspAlaGluGluAsnLysGluAspLysGluLysGluGluGluGluGluLysPro 192
QY 960 -----GGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 1004
Db 193 LysArgGlySerIleGlyGluAsnGlnGlyGluGluLysGlyThrLysValGlnAlaLys 212
QY 1005 AGAACT-----AGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1037
Db 213 ArgGluLysLeuGlnGluAspLysProThrPheLysLysGluGluLysAspGluLys 232
QY 1038 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG--- 1079
Db 233 IleLysLysAspLysGluProLysGluGluValLysSerPheMetAspArgLysLysGly 252
QY 1079 -----CAAAGAGAGAG--- 1079
Db 253 PheThrGluValLysSerGlnAsnGlyGluPheMetThrHisLysLeuLysHisThrGlu 272
QY 1080 -----AAGAAAGAGAGAG--- 1094
Db 273 AsnThrPheSerArgProGlyGlyArgAlaSerValAspThrLysGluAlaGluGlyAla 292
QY 1095 -----AGAACTNAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG--- 1133
Db 293 ProGlnValGluAlaGlyLysArgLeuGluGluLeuArgArgArgGlyGluThrGlu 312
QY 1134 AATAAGGACNAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1175
Db 313 SerGluGluPheGluLysLeuLysGlnLysGlnGlnAlaAlaLeuGluGlu 332
QY 1176 AAGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 1235
Db 333 LeuLysLysArgGluGluArgLysValLeuGluGluGluGlnArgGly 352
QY 1236 GAAAGTAGAAAGCGGAGAGAG--- 1277
Db 352 SGInGluGluAlaAspArgLysLeuArgGluGluGluGluLysArgLeuLysGluGlu 372
QY 1278 AGAAGAGAGAGAGAG 1293
Db 372 uilleGluArgArg 377

RESULT 14
US-09-461-697-188
; Sequence 188, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-188

Alignment Scores:
Pred. No.: 9,72e-13 Length: 257
Score: 209.00 Matches: 77
Percent Similarity: 48.47% Conservative: 50
Best Local Similarity: 29.39% Mismatches: 71
Query Match: 9.77% Indels: 65
DB: 4 Gaps: 11

US-09-502-945-3 (1-1298) x US-09-461-697-188 (1-257)
QY 600 CCAGTGCAGACCCCTTTACTACCCCTGTGAAAGGCTCTCCGAGCCATGAGCCTA 659
Db 4 ProValThr---ProGluValLysProLysArgThrSerSerArgLysMetLysThr 22
QY 660 GAAGAGGCAAGATGCGACGA-----GCAGAGCTTCAG 692
Db 23 LysSerAspMetMetGluGluAsnIleAspThrSerAlaGlnAlaValAlaGluThrLys 42
QY 693 AGGCTCGGGCTCTGACGTCTACTATATGANGCCCAAGGCTCGAAGA----- 737
Db 43 GlnGluAlaValGluGluAspTyrAsnGluAsnAlaLysAsnGlyGluAlaLysIle 62
QY 738 -----GAGAAGAAATCNAAGTTAAAGTATCAAAAGTC----- 773
Db 63 ThrGluAlaProAlaSerGluLysGluIleValGluValLysGluGluAsnIleGluAsp 82
QY 774 ---GTGAGAAAGAGAAAGGCCCAAGAA-----GCCTAAAGAGTTTGAG 815
Db 83 AlaThrGluLysGlyGluLysLysGluAlaValAlaGluValLysAsnGluGlu 102
QY 816 CAGTCGCGAAGGTAAATCCAGCTGCCCTAGAACGCA-----AGAAAGAG 866
Db 103 GluAspGlnLysGluAspGluGluAsnGlnAspGlnGluLysGlyGluAlaGlyLysGlu 122
QY 867 GAAAGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902
Db 123 AspLysAspGluLysGlyGluGluAspGlyLysGluAspLysAsnGlyLysGly 142
QY 903 -----AAGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 926
Db 143 GluAspAlaLysGluLysGluAspGlyLysGlyGluAspGlyLysGlyAsnGlyGlu 162
QY 927 ---AGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 983
Db 163 AspGlyLysGluLysGlyGluAspGluGluGluGluAspArgLysGluThrGlyVal 182
QY 984 GAAAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1043
Db 183 GlyLysGluAsnGluAsp-----GlyLysGluLysGlyAspLysGlyLysGly 198
QY 1044 AGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1099
Db 199 LysAspValLysValLysGluAspGluLysGluAspGlyLysGluAspGlyLys 218
QY 1100 TNAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1159
Db 219 GlyAsnGluGluAlaGlyLysGluLysGluAsp-----LeuLysGluGlu 234
QY 1160 AAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1219
Db 235 GluGluGlyLysGluGluAspGluIleLysGlu---AspAspGlyLysLysGluGluPro 254
QY 1220 AA 1221
Db 254 In 254
```


RESULT 15

US-09-461-697-186
; Sequence 186, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: LO, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 186
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-186

Alignment Scores:

Pred. No.:	9.93e-13	Length:	272
Score:	209.00	Matches:	77
Percent Similarity:	48.47%	Conservative:	50
Best Local Similarity:	29.39%	Mismatches:	71
Query Match:	9.77%	Indels:	65
DB:	4	Gaps:	11

US-09-502-945-3 (1-1298) x US-09-461-697-186 (1-272)

QY	600	CCAGTGACAGACCTTTTACTGACCCCTCTCGGAAAGGCTCTCCGAGCCATGAGCCTA	659
DB	19	ProValThr---ProGluValLysProLysArgThrSerSerArgLysMetLysThr	37
QY	660	GAAGAGGCAAGATCGCGACGA-----GCAGAGCTTCAG	692
DB	38	LysSerAspMetMetGluGluAsnIleAspThrSerAlaGlnAlaValAlaGluThrLys	57
QY	693	AGGCTCGGGCTCTGCAGTCTCTACTATGANGCCAGGCTCGAAGA-----	737
DB	58	GlnGluAlaValAlaGluGluAspTyrAsnGluAsnAlaLysAsnGlyGluAlaLysIle	77
QY	738	-----GAGAAGAAATCNAAAGTTTAAAGTATACAAAGTC-----	773
DB	78	ThrGluAlaProAlaSerGluLysGluIleValGluValLysGluGluAsnIleGluAsp	97
QY	774	---GTGAAGAAAGGAAAGGCCAAGAA-----GCCCTAAAGAGTTTGAG	815
DB	98	AlaThrGluLysGlyGlyGluLysGluAlaValAlaAlaGluValLysAsnGluGlu	117
QY	816	CAGCTGCGGAGGTAAATCCAGCTGCCCTAGCAAGCA-----AGAAAGAG	866
DB	118	GluAspGlnLysGluAspGluGluAspGlnAsnGluGluLysGlyGluAlaGlyLysGlu	137
QY	867	GAAGAAGGAGGAGGAGAAAGAAAGAAACAAGGAG-----	902
DB	138	AspLysAspGluLysGlyGluGluAspGlyLysGluAspLysAsnGlyAsnGluLysGly	157
QY	903	-----AAGAAGAAAGAAAGGAGAGAG-----	926
DB	158	GluAspAlaLysGluLysGluAspGlyLysGlyGluAspGlyLysGlyAsnGlyGlu	177
QY	927	---AGAAGAAAGAGGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	983
DB	178	AspGlyLysGlyGluAspGluGluAspGluGluAspArgLysGluThrGlyVal	197
QY	984	GGAAGAGAGGAG	1043

DB	198	GlyLysGluAsnGluAsp-----GlyLysGluLysGlyAspLysLysGluGly	213
QY	1044	AGAAAGAAAGTAAAGAAAGAAAG-----AAGGAAGAAAGAAAGAAAGAAAG	1099
DB	214	LysAspValLysValLysGluAspGluLysGluArgGluAspGlyLysGluAspGluGly	233
QY	1100	TNAGAAGAAAGAGAGGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGAGAG	1159
DB	234	GlyAsnGluGluAlaGlyLysGluLysGluAsp-----LeuLysGluGlu	249
QY	1160	AAAGAATAAGAGAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	1219
DB	250	GluGluGlyLysGluGluAspGluLysGluAspGluLysGluAspGluLysGlu	269
QY	1220	AA 1221	
DB	269	ln 269	

Search completed: March 21, 2003, 13:08:07
Job time : 21.0536 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 13:00:02 ; Search time 13.8984 Seconds
(without alignments)
9985.861 Million cell updates/sec

Title: US-09-502-945-3
Perfect score: 2140
Sequence: 1 ggcgtcgtgaatgactgcga.....gaagaagaagaagaaaaa 1298

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 442306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-Q/cgn2_1/USPRO_spool/US09502945/runat_14032003_101101_19301/app_query.fasta_1.10979
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09502945 -ECGN_1_1_60 -runat_14032003_101101_19301
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	401	18.7	434	10	US-09-124-280A-3
2	313	14.6	611	10	US-09-216-393-81
3	279	13.0	369	9	US-09-820-843A-95
4	277.5	13.0	213	12	US-10-001-843-178

5	260.5	12.2	197	9	US-10-101-487-51	Sequence 51, Appl
6	256.5	12.0	179	9	US-10-101-487-107	Sequence 107, Appl
7	256.5	12.0	197	9	US-10-101-487-114	Sequence 114, Appl
8	256.5	12.0	1005	10	US-09-925-301-1335	Sequence 1335, Appl
9	255.5	11.9	180	9	US-10-101-487-116	Sequence 116, Appl
10	251.5	11.8	176	9	US-10-101-487-56	Sequence 56, Appl
11	251	11.7	130	10	US-09-864-761-35536	Sequence 35536, A
12	250	11.7	186	9	US-10-101-487-44	Sequence 44, Appl
13	250	11.7	198	9	US-10-101-487-42	Sequence 42, Appl
14	250	11.7	200	9	US-10-101-487-53	Sequence 53, Appl
15	248	11.6	179	9	US-10-101-487-46	Sequence 46, Appl
16	248	11.6	240	9	US-10-101-487-75	Sequence 75, Appl
17	248	11.6	350	9	US-10-101-487-58	Sequence 58, Appl
18	247.5	11.6	376	10	US-09-925-301-1399	Sequence 1399, Appl
19	247	11.5	176	9	US-10-101-487-70	Sequence 70, Appl
20	247	11.5	187	9	US-10-101-487-50	Sequence 50, Appl
21	247	11.5	191	9	US-10-101-487-81	Sequence 81, Appl
22	246	11.5	177	9	US-10-101-487-48	Sequence 48, Appl
23	246	11.5	177	9	US-10-101-487-115	Sequence 115, Appl
24	246	11.5	181	9	US-10-101-487-45	Sequence 45, Appl
25	245.5	11.5	174	9	US-10-101-487-72	Sequence 72, Appl
26	245.5	11.5	175	9	US-10-101-487-57	Sequence 57, Appl
27	240.5	11.2	358	9	US-10-001-835-173	Sequence 173, Appl
28	239	11.2	617	10	US-09-864-761-36182	Sequence 36182, A
29	234.5	11.0	432	10	US-09-840-787-47	Sequence 47, Appl
30	234	10.8	242	10	US-09-864-761-36180	Sequence 36180, A
31	221.5	10.4	87	10	US-09-864-761-39625	Sequence 39625, A
32	220	10.3	665	9	US-09-820-843A-107	Sequence 107, Appl
33	218.5	10.2	89	10	US-09-864-761-35241	Sequence 35241, A
34	217.5	10.2	86	10	US-09-864-761-34671	Sequence 34671, A
35	213	10.0	1175	10	US-09-771-161A-224	Sequence 224, A
36	213	10.0	1175	10	US-09-771-161A-225	Sequence 225, Appl
37	213	10.0	1233	9	US-09-291-417-89	Sequence 89, Appl
38	213	10.0	1233	9	US-09-291-417-13	Sequence 13, Appl
39	213	10.0	1233	9	US-09-291-417-13	Sequence 13, Appl
40	212.5	9.9	452	10	US-09-881-752A-268	Sequence 268, Appl
41	212	9.9	243	9	US-10-001-835-169	Sequence 169, Appl
42	210	9.8	259	9	US-09-925-299-821	Sequence 821, Appl
43	210	9.8	259	10	US-09-925-299-821	Sequence 821, Appl
44	209	9.8	257	10	US-09-922-261-188	Sequence 188, Appl
45	209	9.8	272	10	US-09-922-261-186	Sequence 186, Appl

ALIGNMENTS

RESULT 1
US-09-124-280A-3
Sequence 3, Application US/09124280A
Patent No. US20020034520A1
GENERAL INFORMATION:
APPLICANT: POTRO, Massimo
TITLE OF INVENTION: VACCINES FOR PREVENTION OF GRAM-
NEGATIVE BACTERIAL INFECTIONS AND
NUMBER OF INVENTION: 45
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hedman, Gibson & Costigan
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/124,280A
FILING DATE: July 29,1998
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Costigan, James V.
REGISTRATION NUMBER: 25,669
REFERENCE/DOCKET NUMBER: 576-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 302-8989
TELEFAX: (212) 302-8998
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 434 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-09-124-280A-3

Alignment Scores:		
Pred. No.:	5,81e-26	Length:
Score:	401.00	Matches:
Percent Similarity:	73.33%	Conservative:
Best Local Similarity:	35.11%	Mismatches:
Query Match:	18.74%	Indels:
DB:	10	Gaps:
		4
		0
		434
		79
		86
		56
		5
		4

US-09-502-945-3 (1-1298) x US-09-124-280A-3 (1-434)

[illegible]


```

Db 556 spGluGluArgProArgLeuGlyAspGluAspArgGluProSerLeuArgProA 576
QY 1280 AGAAGAAAGA 1290
Db 576 spAspArg 579
RESULT 9
US-10-101-487-116
; Sequence 116, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOUIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 116
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-116
Alignment Scores:
Pred. No.: 9,49e-14 Length: 180
Score: 255.50 Matches: 65
Percent Similarity: 61.66% Conservative: 54
Best Local Similarity: 33.68% Mismatches: 48
Query Match: 11.94% Indels: 26
DB: 9 Gaps: 1
US-09-502-945-3 (1-1298) x US-10-101-487-116 (1-180)
QY 728 GGCCTCAAGAGAGAAATCNAAAGTTAAAGTATCACAAGTCTGTAAGAAAGAAA 787
Db 1 GlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 20
QY 788 GGCCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 847
Db 21 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 26
QY 848 AGAAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 907
Db 27 -----GluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 39
QY 908 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 967
Db 40 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 59
QY 968 AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1027
Db 59 luGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 78
QY 1028 GAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1087
Db 79 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 98
QY 1088 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1147
Db 99 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 118
QY 1148 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1207

```

```

Db 119 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 138
QY 1208 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1267
Db 138 luGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 158
QY 1268 ATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
Db 158 luGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 167
RESULT 10
US-10-101-487-56
; Sequence 56, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOUIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-56
Alignment Scores:
Pred. No.: 2,06e-13 Length: 176
Score: 251.50 Matches: 65
Percent Similarity: 62.69% Conservative: 56
Best Local Similarity: 33.68% Mismatches: 54
Query Match: 11.75% Indels: 18
DB: 9 Gaps: 1
US-09-502-945-3 (1-1298) x US-10-101-487-56 (1-176)
QY 728 GGCCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 787
Db 1 GlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 20
QY 788 GGCCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 847
Db 21 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 27
QY 848 AGAAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 907
Db 28 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 47
QY 908 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 967
Db 48 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 67
QY 968 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1027
Db 67 luGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 86
QY 1028 GAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1087
Db 87 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 106

```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:32:38 ; Search time 23.1214 Seconds
(without alignments)
10793.685 Million cell updates/sec

Title: US-09-502-945-3

Perfect score: 2140

Sequence: 1 ggctgctgaatgactgcga.....gaagaagaagaagaaaaa 1298

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model
-Q=/cgn2_1/USPTO-spool/US09502945/runat_14032003_101059_19127/app_query.fasta_1.10979
-DB=PIR_73 -FASTA=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_263 -runat_14032003_101059_19127 -NCPU=6 -ICPU=3
-NO_XLPYX -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	322.5	15.1	699	2 E84565	hypothetical prote
2	312.5	14.6	1701	2 T09127	probable erythrocy
3	307.5	14.4	1178	2 S78475	mannosylphosphoryl
4	298	13.9	1898	1 A45973	trichohyalin - hum
5	289.5	13.3	1407	1 S28589	trichohyalin - rab
6	287.5	13.4	335	2 T33457	hypothetical prote
7	285	13.3	1166	2 H86341	hypothetical prote
8	280	13.1	385	2 T19201	hypothetical prote
9	279	13.0	369	2 H71321	conserved hypothet
10	275	12.9	451	2 G70241	hypothetical prote
11	274.5	12.8	390	2 T34137	hypothetical prote
12	272	12.7	647	2 T29523	hypothetical prote
13	269	12.6	1240	2 S52734	hypothetical prote
14	267.5	12.5	1390	2 S51364	sperm tail-specifi

15	266	12.4	1052	1 A44937	kinetoplast-associ
16	264.5	12.4	312	2 T25994	hypothetical prote
17	264.5	12.4	810	2 T44430	protein PV100 [imp
18	262.5	12.3	877	2 T50591	class II INCENP pr
19	259	12.1	1432	2 B85431	trichohyalin like
20	259	12.1	1549	1 A40691	trichohyalin - she
21	258.5	12.1	771	1 A33430	h-caldesmon - chic
22	257.5	12.0	678	2 A54514	glutamic acid-rich
23	255.5	11.9	1085	2 S62516	hypothetical coile
24	254	11.9	899	2 S49634	hypothetical prote
25	253.5	11.8	729	2 T50989	hypothetical prote
26	251	11.7	929	2 T38948	hypothetical coile
27	250.5	11.7	522	2 C96608	hypothetical prote
28	249.5	11.7	849	2 S00030	neurofilament trip
29	249	11.6	1359	2 T34036	hypothetical prote
30	248	11.6	1344	2 T42637	hypothetical prote
31	247.5	11.6	484	2 A40988	54K arginine-rich
32	247.5	11.6	2962	2 T19756	hypothetical prote
33	246.5	11.5	301	2 T33068	hypothetical prote
34	246	11.5	793	1 JH0628	caldesmon - human
35	245	11.4	845	2 A45669	neurofilament trip
36	244	11.4	839	2 T50590	class I INCENP pro
37	243.5	11.4	1403	2 T11583	probable translati
38	242.5	11.3	737	2 T15597	hypothetical prote
39	242.5	11.3	1017	2 T15598	hypothetical prote
40	241.5	11.3	980	2 E71606	hypothetical prote
41	240	11.2	1877	2 T21861	hypothetical prote
42	239	11.2	1020	1 QFHUH	neurofilament trip
43	238.5	11.1	409	2 E86336	hypothetical prote
44	238.5	11.1	447	2 S52391	centrosomin B - mo
45	235.5	11.0	762	2 G88436	protein T04A8.13

ALIGNMENTS

RESULT 1

E84565
hypothetical protein At2g18540 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84565
R:Lin. X.: Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.: Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, C.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; UID:20083487; PMID:10617197
A:Cross-references: GB:AE002093; NID:g4218005; PIDN:AAD12213.1; GSPDB:GN00139
A:Accession: E84565
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-699 <STO>
A:Cross-references: GB:AE002093; NID:g4218005; PIDN:AAD12213.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g18540
A:Map position: 2

Alignment Scores:	Pred. No.:	Length:	Matches:
Score:	322.50	699	119
Percent Similarity:	43.66%	Conservative:	91
Best Local Similarity:	24.74%	Mismatches:	138
Query Match:	15.07%	Indels:	133
DB:	2	Gaps:	17

US-09-502-945-3 (1-1298) x E84565 (1-699)

QY	2	GCTGCTGAATGACTGCGAACCCGCTTCGACAGAGAGCTTCTGCTTTGAGCCANCAAGAA	61
Db	292	AlaCysGluLeuSerIleValLeuGluGlyMetValArgValValAsnGlnGln	311
QY	62	GAAGTACGGATTTGCCAAAGACTACCTCTTACGTGAGAGT-----GAAGATGAGGGG	115
Db	312	SerLeuSerSerCysLysAsnAsp---ArgLysSerGluSerPheMetValGluGluGly	330


```
QY 116 GAC----- 118
|||
Db 331 AspValPheValProLysPheHisProMetAlaGlnMetSerPheGluAsnSerSer 350
QY 119 -----AATGATGGAGAGAGAAAGCATCNAAGCTTCTGGAACA 157
|||
Db 351 PheValPheMetGlyPheSerThrSerAlaLysThrAsnHisProGlnPheLeuValGly 370
QY 158 ATCAAGTCCCTTGATGGAAAGATAGCGGAAATGGCTGAGNAGTCTGAGGCTAGTCTG 217
|||
Db 371 GlnSerSerVal-----LeuLysValLeuAspArgAspValValAlaVal 385
QY 218 AAGGTGTCAGAGTCAATGTCAGTCTCTGAAGGATCAGGAGAAAAGCTGCTTCACAGAT 277
|||
Db 386 -----SerPheAsnLeuSerAsnGlu-----ThrIleLysGly 396
QY 278 CTGCTTGACCTGTAAACTTCA-----TCTTCTTGGCCACTCTGAAAAGCNA 328
|||
Db 397 LeuLeuLysAlaGlnLysGluSerValIlePheGluCysAlaSerCysAlaGluGlyGlu 416
QY 329 CTCAGTAGAGTCAATCAAAAGAACAGTGGAGTTACCTCTGAACAAGAGAGATTGAA 388
|||
Db 417 LeuSerLysLeuMetArgGluIleGluLysArgLysArgGluGluGluIleGlu 436
QY 389 CGGATCCACAGAGATPAGCATTTCAATPAAACGCACACAGTCTCTCCAAATGGGACCTGT 448
|||
Db 437 ArgArgArgLysGluGluGluAlaArgLysArgGluGluAlaLysArg----- 453
QY 449 CGTCTTGAGAACCGCAGCAGCAGCAGTGGTTTCCCTGGAGAAAGAGAGCCAGC 508
|||
Db 454 ArgGluGluGluAlaLysArg----- 462
QY 509 CATGTCTCCATGAACATGTGCTCAGTGTGCTGGAAGCAAGAACCTCCCTGGAGCAGA 568
|||
Db 462 ----- 462
QY 569 AATTTTCAACCTCTCCATGAAGAACAGCAGCAGTGCAGACAGCCCTTTACTACCCCTT 628
|||
Db 463 -----GluGluGluThrGluArgLysLysArgGlu----- 473
QY 629 GGAAGAGCCTCTCTCGAGCCATAGCTAGAGAGGCAAGATCGCAGCAGCAGCT 688
|||
Db 474 -----GluGluGluAlaArgLysArgGluGluGluArgLysArgGlu 487
QY 689 TCAGAGGGCTCGGGCTCTGCAGTCTTACTATGANGCCAAAGCTCGAAGAGAGAAAT 748
|||
Db 488 GluGlu-----GluAlaLysArgArgGluGluGlu 497
QY 749 CNAAGTTAAAGTATCACAAGTCTGTGAAGAAAGGCAAGAAAGCCCTAAAGA 808
|||
Db 498 ArgLysLysArgGluGluGluAlaGluGlnAlaArgLysArgGluGluArgGluLys 517
QY 809 GTTTGAGCAGCTCGGAAGTTAATCCAGCTGCCACCTAGAAAGCAGAAAGAGAGA 868
|||
Db 518 Glu-GluGluMetAlaLys-----LysArgGluGluGlu 528
QY 869 AAGAAGCAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 928
|||
Db 528 uArgGlnArgLysGluGluGluValGluArgLysArgGluGluGluGluGluGluArgLys 548
QY 929 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 979
|||
Db 548 ArgArgGluGluGluAlaArgLysArgGluGluGluArgLysArgGluGluGluMetal 568
QY 980 AAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1033
|||
Db 568 aLysArgArgGluGluGluGluArgGlnArgLysGluArgGluGluValGluArgLysIleArg 588
QY 1034 GAAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075
|||
Db 588 gGluGluGlnGluArgLysArgGluGluGluMetalAlaLysArgArgGluGluGluArgGlu 608
```

```
QY 1076 GAAGAAGAAAGAGAGAGAACTNAGAGAGAAAGAGAGAGAGAGAGAGAGAGAA 1135
|||
Db 608 nLysLysGluArgGluGluMetGluArgLysLysArgGluGluGluAlaArgLysArgGlu 628
QY 1136 TAAGAACACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1186
|||
Db 628 uGluGluMetAlaLysIleArgGluGluGluArgGluGluArgLysGluArgGluAspValGlu 648
QY 1187 GAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
|||
Db 648 uArgLysArgArgGluGluGluAlaMetArgArgGluGluGluArgLysArgGluGluGlu 668
QY 1238 AAGTAGAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
|||
Db 668 uAlaAlaLysArgAlaGluGluGlu-ArgArgLysLysGluGluGluGluGluArg 687
```

RESULT 2

T09127

probable erythrocyte-binding protein MAEBL - Plasmodium yoelii

C:Species: Plasmodium yoelii

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999

C:Accession: T09127

R:Kappe, S.H.I.; Noe, A.R.; Fraser, T.S.; Blair, P.L.; Adams, J.H.

Proc.Natl. Acad. Sci. U.S.A. 95, 1230-1235, 1998

A:Title: A family of chimeric erythrocyte binding proteins of malaria parasites.

A:Reference number: Z16577; MUID:98115903; PMID:9448314

A:Accession: T09127

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1701 <KAP>

A:Cross-references: EMBL:AF031886; NID:G2947227; PID:G2947228

A:Experimental source: subspecies yoelii; strain YM

A:Genetics:

A:Gene: mael

A:Introns: 62/1; 1648/1; 1674/2; 1697/1

C:Keywords: alternative splicing; cell binding; erythrocyte invasion

Alignment Scores:

Pred. No.:	1.01e-14	Length:	1701
Score:	312.50	Matches:	99
Percent Similarity:	52.17%	Conservative:	69
Best Local Similarity:	30.75%	Mismatches:	114
Query Match:	14.60%	Indels:	40
DB:	2	Gaps:	12

US-09-502-945-3 (1-1298) x T09127 (1-1701)

```
QY 417 AAGCAGACAGTCTCTCCAA-----TGGGACCTGTCTC 452
|||
Db 1098 AsnAlaGluIleIleArgLysPheGluGluAlaGlnLysAlaAlaIleAlaLysLysAla 1117
QY 453 CTGAAGAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 512
|||
Db 1118 GluGluGluArgLysLysAlaGluAlaValLysLysAlaGluGluGluArgLysArgIle 1137
QY 513 -----GCTCCCATTTGAACATGCTGCTGAGTGGCTGGAAGGCAAGAACCTCCCTGGAG 563
|||
Db 1138 GluAlaGluLysLysAlaGluGluGluArgLysArgIleGluAlaGluLysLysAlaGlu 1157
QY 564 CAGGAATTTTCAACCTCTCCATGAAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 623
|||
Db 1158 GluGluArgLysArgIleGluAlaGluLysLysAlaGluGluGluArgLysIleIleGlu 1177
QY 624 CCGTGGGAAAGCGCTCTCTCGAGCCATGAGCCCTAGAGAGAGAGAGAGAGAGAGAGAG 683
|||
Db 1178 AlaAlaLysLysAlaGluGluGluArgLysArgIleGluGluAlaLysLysAlaGluGlu 1197
QY 684 GAGCTTCAGAGGCGCTGGGCTCTGCAGCTCTACTACTATGANGCCAGCGCTCGAGAGAGAG 743
|||
Db 1198 GluArgLysLysIleGluAlaLys-----LysAlaGluGluGluArg 1212
QY 744 AAAATCNAAAGTTAAAGATATCATCAAGATGCTGTAAGAGAGAGAGAGAGAGAGAGAG 797
|||
```


Db 1213 Lys-----LysAlaGluAlaValLysLysLysAlaGluGluAlaLysLysLysAla 1228
 QY 798 ---GCCCTAAAGAGCTTTGACGAGCTGCGGAGGTTAATCCAGCTGCC-----CCACTA 848
 Db 1229 GluAlaLysLysAlaGluGluArgLysLysLysLysAlaGluAlaLysLysLysAlaLeu 1248
 QY 849 GAAGACCAAGAAAG 908
 Db 1249 GluArgLysLysLysSerGlu-----AlaAlaLysLysAlaLeuGluGluArgLysLysLys 1266
 QY 909 AAAGAAG 968
 Db 1267 AlaGluAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1286
 QY 969 AAAGGAG 1028
 Db 1286 uGluGluLysLysLysAlaGluAlaLysLysLysLysLysLysLysLysLysLysLys 1306
 QY 1029 AATAAG 1085
 Db 1306 sLeuGluArgLysProGluArgLysLysLysLysLysLysLysLysLysLysLysLys 1326
 QY 1086 AGAAG 1137
 Db 1326 gLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1346
 QY 1138 ---AGAACNAG 1193
 Db 1346 gGluArgLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1366
 QY 1194 AGAGAA-----GGAAG 1232
 Db 1366 sGluLysLysAlaGluAlaLysLysLysLysLysLysLysLysLysLysLysLysLys 1386
 QY 1233 GAAGAAAGTAGAAGCGGAG 1292
 Db 1386 uLysLysAlaGluGluGluArgLysLysLysLysLysLysLysLysLysLysLysLys 1406
 QY 1293 GAAA 1296
 Db 1406 gLys 1407
 RESULT 3
 S78475
 N;Alternate names: protein MNN4 - yeast (Saccharomyces cerevisiae)
 C;Species: Saccharomyces cerevisiae
 C;Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 19-Apr-2002
 C;Accession: S78475; S38037; S38038
 R;Odani, T.; Shimizu, Y.; Yoshifumi, J.
 submitted to the EMBL Data Library, January 1996
 A;Description: Cloning and Analysis of the MNN4 Gene Required for Phosphorylation of N-1
 A;Reference number: S78475
 A;Accession: S78475
 A;Molecule type: DNA
 A;Residues: 1-1178 <ODA>
 A;Cross-references: EMBL:D83006; NID:g1752735; PID:d1012343; PID:g1752736; MIPS:YKL201c
 A;Note: this is a revision to the sequence from reference S38024
 R;Maia e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.; Guerreiro, P.; H
 submitted to the Protein Sequence Database, March 1994
 A;Reference number: S38024
 A;Accession: S38037
 A;Molecule type: DNA
 A;Residues: 121-249, 'ATIQLOT', 255, 'MALLRD', 262-390, 'LRISSEN', 398-515, 'LG' <MAI>
 A;Cross-references: EMBL:D28200; NID:g486355; PID:g486356
 A;Experimental source: strain S288C
 A;Note: this sequence has been revised in reference S78475
 A;Note: this was assumed to be protein YKL200c
 A;Accession: S38038
 A;Molecule type: DNA
 A;Residues: 640-1178 <MAW>
 A;Cross-references: EMBL:D28201; NID:g486357; PID:g486359; MIPS:YKL201c
 A;Experimental source: strain S288C

A;Note: this sequence has been revised in reference S78475
 A;Note: this was assumed to be the complete sequence of protein YKL201c
 C;Genetics:
 A;Gene: SGD:MNN4; MNN4
 A;Cross-references: MIPS:YKL201c; SGD:S0001684
 A;Map position: 11L
 C;Function:
 A;Description: required for phosphorylation of N-linked oligosaccharides
 C;Keywords: transmembrane protein
 F;28-44/Domain: transmembrane #status predicted <TM>

Alignment Scores: 2,3e-14 Length: 1178
 Pred. No.: 307/50 Matches: 69
 Score: 69.94% Conservative: 52
 Percent Similarity: 39.88% Mismatches: 37
 Best Local Similarity: 14.37% Indels: 15
 Query Match: 2 Gaps: 4
 DB:

US-09-502-945-3 (1-1298) x S78475 (1-1178)
 QY 792 AGAAGCCCTAAAGAGTTTGACGAGCTGCGGAGGTTAATCCAGCTGCCGACTAGAA 851
 Db 1011 ArgLysGlyIleGlnMetPheAspLys-----AspProIleIleValTyrGlu 1026
 QY 852 GAA-----CCAAGAAAAG 899
 Db 1027 AspTyrAlaTyrAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1046
 QY 900 GAGAAGAAAGAAAG 959
 Db 1047 GluGluGlu-LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1066
 QY 960 GGAGAAAGAAAG 1019
 Db 1066 sLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1085
 QY 1020 -----GAGAGGAG 1073
 Db 1086 GluGluGluGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1105
 QY 1074 AGGAAG 1133
 Db 1106 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1124
 QY 1134 AATAAGACACNAG 1193
 Db 1125 AsnGluAspGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1144
 QY 1194 AGAAG 1253
 Db 1145 AsnLysLysAsnGluAspGluGluLysLysLysLysLysLysLysLysLysLysLys 1164
 QY 1254 AAAGAAAGAAAGTATAAG 1284
 Db 1165 GluGluGlu-GluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1174

RESULT 4

A45973
 trichohyalin - human
 C;Species: Homo sapiens (man)
 C;Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
 C;Accession: A45973
 R;Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
 J. Biol. Chem. 268, 12164-12176, 1993
 A;Title: The structure of human trichohyalin. Potential multiple roles as a functional
 ed (cross-linking) protein.
 A;Reference number: A45973; MUID:93280194; PMID:7685034
 A;Accession: A45973
 A;Molecule type: DNA
 A;Residues: 1-1898 <LEE>
 A;Cross-references: GB:L09190; NID:g292835; PIDN:AAA65582.1; PID:g292836
 A;Note: authors translated the codon AGG for residue 1714 as Pro

C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sheath. Covalent modifications to this protein include conversion of arginine to citrulline and covalent modifications to citrulline.

C;Genetics:

A;Gene: GDB:THH

A;Cross-references: GDB:136223; OMIM:190370

A;Map position: 1q21-1q21

C;Superfamily: trichohyalin; calmodulin repeat homology

C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F;49-81/Domain: calmodulin repeat homology <EF2>

Alignment Scores:

Pred. No.: 1.09e-13 Length: 1898
Score: 298.00 Matches: 96
Percent Similarity: 48.10% Conservativeness: 106
Best Local Similarity: 22.86% Mismatches: 136
Query Match: 13.93% Indels: 82
DB: 12 Gaps: 12

US-09-502-945-3 (1-1298) x A45973 (1-1898)

```

QY 77 CCAAAAGACTACCTCTGAGTGAGAGTGAAGTGGGGGAGCAATGATGAGAGAGAAAG 136
Db 124 ProArgAspArgGlnLeuGluGlu-----GluProGlyGlnArgArg 137
QY 137 CATCNAAAGCTTCTGGAACCAATCAGTTCCTTCATGGAAGAAATAGCGGAAATTTGGCT 196
Db 138 ArgGlnLysArgGlnGlu-----GlnGluArgGluLeuAla 149
QY 197 GANAGGTCTGAGGTAGTCTGAAGGTGTGAGAGTTCAGTTCAGTTCGAAGGATCAGGA 256
Db 150 GluGlyGluGlnSerGluLysGlnGlnGluArgLeuGlnGlnArgAspArgGlnArgArg 169
QY 257 GAAAGCTGGCTTCGAGATCTGCTGAGCTGTAAACTTCATCTTTCGCCACT 316
Db 170 AspGluGluLeuTrpArgGlnArgGlnGluTrpGlnGluArgGluGluArgAlaGlu 189
QY 317 GTGAAAGCACTCAGTGTAGTGCANATCAAGAAAGACAGTGGAGTTCACCT----- 367
Db 190 ---GluGlnGlnSerCysLysGlyHisGluThrGluGluPheProAspGluGlu 208
QY 368 ---CTGACAAAGAGAGATTGAACGGATCCACAGAGATAGCATTCATTAACGCCACA 424
Db 209 GlnLeuArgArgGluLeuLeuGlnGluLeuArgArgLysGlyArgGluGluGlnGln-Gl 228
QY 425 AGTCCTCTCCAAATGGACCCCTGCTGCTGTGAAGACCGGAGGAGGAGGAGGAGGAGG 484
Db 228 n-----GlnArgArgGluArgGlnArgGlnArgGlnArgGlnArgGlnArgGln 238
QY 485 TCCCTGGAGAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 544
Db 238 eGlnGluGluGluGluGluGlu-----ThrValLeuArgLysGluGluGluLysLe 247
QY 545 GGCAGAGAACTCCCTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 604
Db 247 GlyArgGlu-----ThrValLeuArgLysGluGluGluLysLe 260
QY 605 GACAGACCTTTACTACCCCTGTGGAAGGCTCTCCAGGAGGAGGAGGAGGAGGAGGAGGAG 664
Db 260 uGlnGluGlu-----GluProGlnArgGlnArgGlnGlnGlnGlnGlnGlnGlnGln 278
QY 665 GGCAGAGATCGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 724
Db 278 uArgLysLeuGluGluGlnGluGluGluGluGluGluGluGluGluGlnGlnGlnGln 298
QY 725 CAAGGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 784
Db 298 nArgLeuArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 308
QY 785 AAAGGCCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 844
Db 309 -----LysGlnGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 316
QY 845 ACTAGAAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 904

```

```

Db 317 -GlnGlnGluGluArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 336
QY 905 GAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 964
Db 336 luGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 356
QY 965 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021
Db 356 luGluGluArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 376
QY 1022 GAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081
Db 376 InLeuArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 396
QY 1082 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
Db 396 euArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 416
QY 1142 ACNAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1177
Db 416 rgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 436
QY 1178 GAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
Db 436 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 456
QY 1238 AAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1293
Db 456 ySHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 474

```

RESULT 5

S28589

trichohyalin - rabbit

C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999

C;Accession: S28589

R;Fietz, M.J.; Rogers, G.E.

A;Submitted to the EMBL Data Library, December 1992

A;Description: Examination of the gene encoding rabbit trichohyalin.

A;Reference number: S28589

A;Accession: S28589

A;Molecule type: DNA

A;Residues: 1-1407 <PIE>

A;Cross-references: EMBL:Z19092; NID:q1746; PIDN:CAA79519.1; PID:g1747

C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root

Covalent modifications to this protein include conversion of arginine to citrulline a

C;Genetics:

A;Introns: 46/3

C;Superfamily: trichohyalin; calmodulin repeat homology

C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F;49-81/Domain: calmodulin repeat homology <EF2>

Alignment Scores:

Pred. No.: 4.42e-13 Length: 1407
Score: 289.50 Matches: 130
Percent Similarity: 44.04% Conservativeness: 99
Best Local Similarity: 25.00% Mismatches: 160
Query Match: 13.53% Indels: 131
DB: 12 Gaps: 20

US-09-502-945-3 (1-1298) x S28589 (1-1407)

```

QY 23 CGGCTTGACAGAGAGCTTCTGCTTTGAGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 82
Db 728 ArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 745
QY 83 GACTACCTCTTGTAGTGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 133
Db 746 GluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 765
QY 134 -----AAGCATCAAGAGCTTCTGGAAGCAATCAGTTCCTTGTGGAAGAGAT 181

```


Db 766 LeuArgArgGluArgAspArgLysPheArgGluGluGluGlnLeuLeuGlnGluArgGlu 785
Qy 182 AGCGGAAATTGGCTGANAAGGCTCAGGCTAGCTGAGGTTG---TCAGAGTTCAATGTC 238
Db 786 GluGluArgLeu---ArgArgGlnGluArgGluArgLysLeuArgGluGluGlnLeu 804
Qy 239 AGTTCTGAAGGATCAGGAGAAAGCTGGTCTCTGAGCTGCTGAGCCTGTTAAACT 298
Db 805 LeuGlnGluArgGluGluArgLeu----- 813
Qy 299 TCATCTCTTTGGCCACCTGTGAAAAGCACTAGTAGTACATCAATCAAGAACACAGT 358
Db 814 -----ArgArgGlnGluArgGluArgLysLeuArgGluGlu 825
Qy 359 GAGTTACCTCTGAACAAGAGAGATTGAACGGATCCACAGAGATACATCAATAAAA 418
Db 826 GluGlnLeuLeuGlnGluArgGluGluGluArgLeuArgGln-----Glu 841
Qy 419 CGCACAACTCTCTCAAAATGGACCCCTGCTGCTGAAGAACCGCAGGAGAG----- 473
Db 842 ArgGluArgLysLeu-ArgGluGluGlnLeuLeuArgGlnGluGlnGluLeuArg 861
Qy 474 ---CAGCTGCTTTCCCTCGGAGAAAGAGAGAGCCACCATGCTCCCATGTGACATGT 529
Db 861 glnGluArgAlaArgLysLeuArgGluGluGlnLeuLeuArgGlnGluGlnGln 881
Qy 530 GCTCAGTGGCTGGAGGCAAGAACTCCCTCGAGCAGGAGAAATTTCAACCTCCCTCATAA 589
Db 881 uLeuArgGlnGluArgAspArgLysLeuArgGluGlu-----GlnLeuLeuArgGln 899
Qy 590 GAACAAGCAGCAGT-----ACAGACCCCTTTACT 619
Db 899 nGluGlnGlnLeuArgGlnGluArgAspArgLysLeuArgGluGluGlnLeuLe 919
Qy 620 GACCCCTGGAAGAGGCTCTCTC-----CGAGCCATGAGCCTTGA 661
Db 919 uGlnGluSerGluGluArgGlnGluGluGlnLeuLeuArgGlnGluArgLysLeuArgGln 939
Qy 662 AGAGCAAGATGCGACGA-----GCAGAGCTCAGAGGCTCGGGCTCGCAGTCTCTA 715
Db 939 uGlnLeuLeuArgArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 955
Qy 716 CTATGANGCCAAAGCTCGAGAGAGAGAAATCAAAGTTAAAGTATCAAAAGTCTGT 775
Db 956 -----ArgLysLeuArgGluGluGlnLeuLeuGlnGluArgGluGluArgLe 973
Qy 776 GAAGAAGGAAGCCAAAGAGCCCTTAAAGATTTGAGCAGCTG----- 821
Db 973 uArgGlnGlnGluArgAlaArgLysLeuArgGluGluGlnGlnLeuLeuArgGlnGln 993
Qy 822 -----CGGAAGTTAATCCACCTGCCGCTAGAGACG 856
Db 993 uGlnGluLeuArgGlnGluArgAspArgLysPheArgGluGlnGlnLeuLeuGlnGln 1013
Qy 857 AAGAAAGAGGAAGAGAGGAGG----- 881
Db 1013 uArgGluGluArgGlnGluArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1033
Qy 882 -----AGAAGAGAGAACAGAGAGAGAA 907
Db 1033 glnLeuArgArgGlnGlnLeuGluGlnGlnPheArgGlnGluArgAspArgLysPheAr 1053
Qy 908 GAAAGAGAGGAG 963
Db 1053 gLeuGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1073
Qy 964 -----AAGAAAG 1012
Db 1073 spArgLysPheArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1093
Qy 1013 GAAGAAGAGAGAGAG----- 1026
Db 1093 rgArgGluArgAspArgLysPheArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1113

Qy 1027 --AGAATAAGAGGAAG 1084
Db 1113 luArgLeuArgArgGlnGluArgAlaArgLysLeuArgGluGluGlnGluGlnLeuArg 1133
Qy 1085 AAGAGAGAGAACTNAG 1138
Db 1133 rgGluGlnGlnLeuLeuArgGlnGluArgAspArgLysPheArgGlnGluGlnLeu 1153
Qy 1139 GGAACNAG 1192
Db 1153 euGlnGluSerGlu---GluGluArgLeuArgArgGlnGlnGluArgLysLeuArg 1172
Qy 1193 AAG 1252
Db 1172 luGluGlnGlnLeuLeuGlnGluArgGluGluGlnGluArgGlnGlnGlnGln 1192
Qy 1253 GAAAG 1296
Db 1192 rgLysLeuArgGluGlnGlnLeuLeuArgGlnGlnGlnGlnGlnGlnGlnGln 1210
RESULT 6
T33457
hypothetical protein F36H12.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33457
R:Blanchard, M.; Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F36H12.
A:Reference number: 221346
A:Accession: T33457
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-335 <BLA>
A:Cross-references: EMBL:AF078790; PIDN:AAC26930.1; GSPDB:GNO0022; CESP:F36H12.3
A:Experimental source: strain Bristol N2; clone F36H12
C:Genetics:
A:Gene: CRSP:F36H12.3
A:Map position: 4
A:Introns: 32/3; 227/1; 270/2
Alignment Scores:
Pred. No.: 6,18e-13 Length: 335
Score: 287.50 Matches: 70
Percent Similarity: 64.21% Conservative: 52
Best Local Similarity: 36.84% Mismatches: 48
Query Match: 13.43% Indels: 20
DB: 2 Gaps: 3

US-09-502-945-3 (1-1298) x T33457 (1-335)

Qy 740 GAAGAAATCNAAGTTTAAAGTATCAAAAGTGTGAAGAGAGAGAGAGAGAGAGAG 799
Db 51 GluGluLysGluLysSerLysLysSerLysLysGluLysGluProLysLysGluGlu 70
Qy 800 CCTAAAGAGTTTGAGCAGCTGCGGAAGGTTAATCCAGCTGCCCCACTAGAGAACGAG 859
Db 71 LysLysGlu-----LysSerLysLys 77
Qy 860 AAAAG 919
Db 78 SerGluGluLysLysSerLysLysGluGluLysLysLysLysLysLysLysLysLys 96
Qy 920 GAGAAGGAG 979
Db 97 ---Lys-GluAspLysLysGluGluLysLysGluGluLysLysGluAspLysLys 115
Qy 980 AAAAG 1039
Db 115 spGluLysLysAspGluLysLysAspGluLysLysLysLysLysLysLysLysLys 134
Qy 1040 AAG 1093

[illegible]

RESULT 7

H86341
hypothetical protein F9H16.10 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: H86341
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: AB6141; PMID:21016719; PMID:11130712
A;Accession: H86341
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1166 <STO>
A;Cross-references: GB:AE005172; NID:94836896; PIDN:AAD30599.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Alignment Scores:		
Pred. No.:	9,25e-13	1166
Score:	285.00	93
Percent Similarity:	50.87%	Conservative: 54
Best Local Similarity:	32.18%	Mismatches: 62
Query Match:	13.32%	Indels: 80
DB:	2	Gaps: 11
US-09-502-945-3 (1-1298)	x H863341	(1-1166)

Qy	627	GTGGAAGGCGCTCTCTCCGAGCCATGAGCCTA-----GAGAGGCAAGATG	674
Db	3	ValGluLysSerLysTyArgSerGluAspLeuAspValValGluGluGluAspLeu	22
Qy	675	CGACGACGAGCTTCAGAGCGCTCGGGCTCGAG-----TCCCTACTAT	719
Db	23	LysLysSerArgArgAspArgAspArgSerAsnGluArgLysLysAspLysGly-SerCl	42
Qy	720	GANGCCAAAGGCTCGAAGAGAGAGAAATCNAAGTTTAAAGTATCACAAGT-----	772
Db	42	uLysArgGluGluLysAspArgLysLysArgValLysSer-SerAspSerGluAspA	62
Qy	773	-----CGTCAAGAAAGGAAAGCCCAAGAGCCCTAAAA	806
Db	62	spTyAspArgAspAspAspCluArgGluLysArgLysGluLysGluArgGluArgA	82
Qy	807	GAGTTTGACAGCTGCGGAAGTTTATCCAGCTGCCGCCTAGAGAACAAGAAAGAG	866
Db	82	rg-----ArgArgAspLysAspArgV-----	89

[illegible]

RESULT 8

Tl9201
 hypothetical protein Cllg6.3 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: Tl9201
 R:Wilkinson, J.
 submitted to the EMBL Data Library, March 1996
 A:Reference number: Z19089
 A:Accession: Tl9201
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-385 <WIL>
 A:Cross-references: EMBL:Z70204; PIDN:CAA94113.1; GSPDB:GN000028; CESP:C11G6.3
 A:Experimental source: clone C11G6
 C:Genetics:
 A:Gene: CESP:C11G6.3
 A:Map position: X
 A:Introns: 12/1; 106/3; 171/2; 335/2

Alignment Scores:	2.12e-12	Length:	385
Pred. No.:	Score:	Matches:	84
	280.00	Conservative:	62
Percent Similarity:	46.79%	Mismatches:	97
Best Local Similarity:	26.92%	Indels:	69
Query Match:	13.08%	Gaps:	8
DB:	2		

US-09-502-945-3 (1-1298) x T19201 (1-385)

QY 413 ATAAACGGCACAAAGTCCTCTCCAAATGGGACCCCTGTCGTCTGAAGAACCGGCAGGCA 472

Db	21	IleLysPheAsnProAsnLeuGlnThrProLysAlaGluProGluProMetArgPro	40
Qy	473	GCAGCTGGTTTTCCCTCGGAGAAAGAGCAGCCACATGTGTCCTCAATGAACATGTGCT	532
Db	41	AspThrThrSerGlySerArgProAlaSerSerLysSerAsn-----	56
Qy	533	CAGTGGCTGGAAGCAAGAACTCCCCTGGAGCAGGAAAATTTCACCTCCTCCA-----	586
Db	57	-----TyrHisAsnGluProProProProAlaAsnProProLeuLys	72
Qy	587	-----TAAACAACAGCACGCC	601
Db	73	PheargPheLysAsnLeuPheAsnLeuGlyAspGluAspValLysGluGluProSer	92
Qy	602	AGTGACAGACCCTTTACTGACCCTGTGGAAGAAGCCTCTCTCCGAGCCATGAGCCTAGA	661
Db	93	SerMetThrProGluSerSerArgProGlySerSerLeuGlu-----	106
Qy	662	AGAGGCAAAGATCGGACGACCATTCAGAGGGCTCGGGCTGTGCAGTCTCTACTATGA	721
Db	107	-----ThrProSerSerSerSerSerLysHis-HisHisHisHisH1	121
Qy	722	NGCCAAGGCTCGAAGACAGAGAAAAATCNAAAGTTAAAAATGATCATCAAGTCGTCAACAA	781
Db	121	sLysLysGluArgLysAspLysglu-----HisLysLysHisLysLysL	135
Qy	782	AGGAAGGCCAAGAAAGCCCTAAAAGAGTTTCAGCAGCTCGCGAAGGTTTAATCCAGCTGC	841
Db	135	sAspArgGluHisArg-----	140
Qy	842	CGCAGTAGAAGACGAGAAAGAACAGAAAGGAGGAGGAGAAAGAGAAAGAAAGAAAGGA	901
Db	141	-AspArgGluLysGluArgGluArgAspGluArgLysGluArgGluArgGlnGlnLysGl	160
Qy	902	GAAAGAAAGAAAGAGGAGAGAGAGAGAAAGAAAGAGGAGAGAGAAAGAGGAGAGAGG	961
Db	160	uLysGluArgGluAspAlaAlaArgarg-----GluLeGluGluLysAl	175
Qy	962	AGAAAGAAAGAGAGAGGAGAAAGAAAGAAAGAGAGAGAGAAAGAAAGAAAGAAAGGA	1021
Db	175	aGluMetaspAlaLysArgValalagluGluGluArgLysGluLysGluLysGluLys	195
Qy	1022	GAGGAGAGATNAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	1081
Db	195	sArgArgGluLysLysLysGlnLysGlnLysGlnLysGlnLysGlnLysGluArgSerGluAr	215
Qy	1082	GGAAGAGAGGAGGAACCTNAGAAGAGAAAGACAGGAGGAGAAAGAAAGACAGAAATAGGA	1141
Db	215	gLysGluLysGluArgGluLeuGluArgGlnLysGlnLysSerArgLysGluArgGlu	235
Qy	1142	ACNAGAAAGAGGAGAGAAAGAAATAAGAAAGAGGAGAGAAAGAAAGAAAGAAAGAAAG	1201
Db	235	uLysGluArgGluLysGluArgGluLysGluArgGlnLysGluArgGlnLysGluArgGlnLys	253
Qy	1202	GAAAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	1261
Db	253	nLysGluArgGluLysGluArgGluLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys	272
Qy	1262	GAAAGATTAAGAGGAGAGAAAGAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	1295
Db	273	-----GluAlaArgArgLysLysGluGlu	280

RESULT 9
H71321
conserved hypothetical protein TP0470 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: H71321
R:Fraser, C.M.; Norris, S. J.; Weinstock, G. M.; White, O.; Sutton, G. G.; Dodson
rson, J.; Khalak, H.; Richardson, D.; Howell, J. K.; Chidambaram, M.; Utterb
they, L.; Weidman, J.; Smith, H. O.; Venter, J. C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spiro

A:Reference number: A71250; MUID:983327770; PMID:9665876
A:Accession: H71321
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-369 <COL>
A:Cross-references: GB:AE001223; GB:AE000520; NID:g3322745; PIDN:AAC65451.1; PID:g3322745
A:Experimental source: strain Nicholas
C:Genetics:
A:Gene: TP0470

Alignment Scores:		
Pred. No.:	2.49e-12	Length: 369
Score:	279.00	Matches: 96
Percent Similarity:	48.43%	Conservative: 58
Best Local Similarity:	30.19%	Mismatches: 97
Query Match:	13.04%	Indels: 68
DB:	2	Gaps: 14

Alignment Scores:		2.49e-12		369	
Pred. No.:	Score:	Length:	Matches:	Mismatches:	Indels:
Percent Similarity:	48.43%	279.00	96	97	14
Best Local Similarity:	30.19%	Conservative:	58	68	
Query Match:	13.04%	Mismatches:	58	68	
DB:	2	Gaps:	14		
US-09-502-945-3 (1-1298) x H71321 (1-369)					
QY	450	GTCTGTGAAGAACCGGAGAGCAGAGCTGGTTTTTCCCTCGAGAAA-----GAG	500		
DB	48	LeuLeuSertLeuGlnAlaGlnAlaIleGlyProLeuHisLysAlaAlaGlnGln	67		
QY	501	GAGCCAGCC-----ATTGCTCCCATTAACATGTGCTAGTGCC---539			
DB	68	LysProAlaHisProLysAlaAlaLeuTyrLeuGlyMetAlaTyrLeuGlnThrGlyArg	87		
QY	540	-----TGGAGGGCAAGAACTCCCTGGAGCAGAA-----569			
DB	88	TyrThrGlnAlaIleGlnTrpLeuGlnAsnProProValHisSerGlnGlnTyrAlaHis	107		
QY	570	-----ATTTCACCTCTCCATAAGAACCAAGCAGCCAGTG	605		
DB	108	LeuTyrAlaTyrAsnLeuGlyAsnValTyrPreValGlnHisArgTyrGluGlu	125		
QY	606	ACAGACCCCTTACTGACCCCTGTGGAAAAGGCTCTCTCGAGCCATGAGCCTAGAAGAG	665		
DB	126	-----AlaGlnHisAlaTyrGluGlnAlaLeuAlaLeuLysHis	138		
QY	666	-----GCAAGATGCGACGAGCAGAGCTTCAGAGGGCTCGGCCTGCGAGTCC	713		
DB	139	AspTyrProProAlaLeuLeuAsnArgAlaAsnThrAlaMetLysArg-----GlnAla	156		
QY	714	TACTATGANGCCAGGCTCGAAGAGAGAAGAAATCAAAAGTTAAAGTATCACAAAGTC	773		
DB	157	TyrAlaHisAlaLeuAla-----AspTyrLysLysTyr	167		
QY	774	GTGAAGAAGGAAAGGCCAAGAACCCCTAAAGAGCTTTGACAGCTCGCGAAGGTTAAT	833		
DB	168	ValSerGlnAsnProThr-----AlaSerGlnHisTyrGluValGlnArgMetIle---	184		
QY	834	CCAGCTGCCCGCTACGAAGAA-----CCAAGAAAGAGAGAAAGAGGAGGAGGAGAAAG	887		
DB	185	-----AlaAlaLeuGluGlnTrpLeuGlnArgLysGluAlaGluGluAlaArgArgLys	202		
QY	888	AAAGAAACCAAGGAAGAAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	947		
DB	203	GluAlaGluAlaArgArgLysGluAlaGluGluAlaArgArgLysGluAlaGluGlu	222		
QY	948	-----GAAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1004		
DB	223	AlaArgArgLysGluAlaGlu-----GluAlaArgArgLysGluAlaGluGluAla	239		
QY	1005	AGAACTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1064		
DB	240	ArgArgLysGluAlaGluGluAlaArgArgLysGluAlaGluGluAlaArgArgLysGlu	259		
QY	1065	GAAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1123		
DB	260	AlaGlu-GluAlaArgArgLysGluAlaGluAlaArgArgLysGluAlaGluGluAla	279		
QY	1124	AGAAAGAGAGATAAGGAACNAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1183		

Db 279 aArgArgLysGluAlaGluGluAlaAaGArgGlySGluAlaGluGluAlaAaGArgGlySGl 299
 QY 1184 AAAGAAGAAGAAAGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGTAG 1243
 : : : : : ||||| ||| : : : : : |||||
 Db 299 uAlaGluGluAlaAaGArgGlySGluAlaGluGluAlaAaGArgGlySGluAlaGluGluAl 319
 : : : : : ||||| ||| : : : : : |||||
 QY 1244 AAACGGAAGAAAGAAAGAAAGTATAGAGGAGGAAAGGAAAGGAAAGGAA 1295
 : : : : : ||||| : : : : : ||||| ||||| : : : : : |||||
 Db 319 aArgArgLysGluAlaGlu-----GluAlaAaGArgGlySGluAlaGlu 333
 RESULT 10
 G70241
 hypothetical protein BB116 - Lyme disease spirochete plasmid I/Ip28-4
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: G70241
 R:Fraser, C.M.; Castles, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White-
 son, D.; Peterson, J.; Kerravage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt-
 : Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403685
 A:Accession: G70241
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1451 <KLE>
 A:Cross-references: GB:AF000789; NID:g2690079; PIDN:AAC66203.1; PID:g2690100; TIGR:BB116
 A:Experimental source: strain B31
 C:Genetics:
 A:Genome: plasmid

```

Alignment Scores:
Pred. No.: 4,8e-12 Length: 451
Score: 275.00 Matches: 76
Percent Similarity: 60.96% Conservative: 77
Best Local Similarity: 30.28% Mismatches: 70
Query Match: 12.85% Indels: 28
DB: 2 Gaps: 7

US-09-502-945-3 (1-1298) x G70241 (1-451)

Qy 573 TTCACCTCCCTCCATAAGAAC---AAGCAGCAGTGACAGACCCCTTTACTGACCCCTGTG 629
|||||:|||||:|||||:|||||
Db 21 PheasIIeasGlnLysAspIleLysTyrPro-----ProThr 33
|||||:|||||:|||||:|||||

Qy 630 GAAAAGGCCTCTCTCCGAGCCATGAGCTAGAAAGAGGCAAAAGATCGCAGCAGCAGAGCTT 689
|||||:|||||:|||||:|||||
Db 34 GluLysSerArgProLysThr-----GluSerSerLysGlnLysLysLysPro 50
|||||:|||||:|||||:|||||

Qy 690 CAGAGGGCTCGGGCTCTCGAGTCCTACTGATGAGCCAGAGGTCGAGAGAGAGAAGAAATC 749
|||:|||||:|||||:|||||
Db 51 LysThrGluGluGluLeuLys-----LysLysGlnGlnGluGluGluLeu 65
|||||:|||||:|||||:|||||

Qy 750 NAAAGTTAAAAGTATCACAAAGTCGTGAGAAAGCAAGGGCCAGAAAGCCCTAAAA--- 806
|||||:|||||:|||||:|||||
Db 66 LysLysLysGlnGlnGluGluLeuLysLysLysGlnGlnGlnGluGluGluLysLys 85
|||||:|||||:|||||:|||||

Qy 807 -----GAGTTTGACACGCTCGGAAGGTTAATCCAGCTGCGCACTAGAGAAGCAAGA 860
|||:|||||:|||||:|||||
Db 86 LysGlnGlnGluGluGluLeuLysLysLysGlnGlnGlnGluGluGluLeuLysLysGln 105
|||||:|||||:|||||:|||||

Qy 861 AAAGAGGAAAGAGGAGGAGGAGAAAGAAAGCAAGAGGAGAGAAAGAAAGAGAGGG 920
|||||:|||||:|||||:|||||
Db 106 GlnGluGluGluLeuLysLysLysGlnGlnGlnGluGluGluLeuLysLysLysGlnGlnGlu 125
|||||:|||||:|||||:|||||

Qy 921 AGAGGAGAGAAGAAGAGAGAGAGAGAAAGAGGAGAGGAGAGAGAGAGAGAGGA 980
|||||:|||||:|||||:|||||
Db 126 GluGluLeuLysLysLysGlnGlnGluGlu-LeuLysLysLysGlnGlnGluGlu 145
|||||:|||||:|||||:|||||

Qy 981 AAAGGAAAGAGGAGAGAAAGAAAGAACTAGAAAGAGGAGAGGAGAGAAATAAGAGGAA 1040
|||:|||||:|||||:|||||

```


[illegible]

Mol. Biochem. Parasitol. 40, 233-244, 1990

A:Title: Cloning and characterization of a gene coding for a protein (KAP) associated with A:Reference number: A44937; MUID:90301144; PMID:1694571

A:Accession: A44937

A:Molecule type: DNA

A:Residues: 1-1052 <CON>

A:Cross-references: EMBL:M25364; NID:g162141; PIDN:AAA30209.1; PID:g162142

C:Comment: This protein was detected only in kinetoplasts of replicative stages of the P1 insectoplast.

C:Superfamily: kinetoplast-associated protein

C:Keywords: duplication; mitochondrion; tandem repeat

F:424-563,582-705,724-862/Region: 9-residue repeats (A-A-R-K-Q-A-E-E-E)

Alignment Scores:

Pred. No.:	2,09e-11	Length:	1052
Score:	266.00	Matches:	109
Percent Similarity:	40.48%	Conservative:	78
Best Local Similarity:	23.59%	Mismatches:	186
Query Match:	12.43%	Indels:	89
DB:	1	Gaps:	13

US-09-502-945-3 (1-1298) x A44937 (1-1052)

QY	119	AATGATGGAGAGAAAGCATCAATCAAGCTTCTGGAAGCAATCAGTTCCTT-----GAT	172
Db	284	SerSerGluGluAArgHisGluGlnAlaGluLysPheSerGluAlaProLysGlu	303
QY	173	GGAAGAATAGCGGAATTCGGCTGANNAGTCTCAG-----	208
Db	304	AspValLysArgAsnProPheAlaMetMetSerLysLeuMetLysHisSerArgHisPhe	323
QY	209	GCTAGTCTGAAGGTCAGAGTTCATGTCAGTTCTCTGAAGGATCAGGA-----	256
Db	324	GlnThrCysLysThrAlaAlaGluSerProSerAsnThrGlyGluAsnAlaGlnThrGln	343
QY	257	-----GAAAGACTGGCTTCGAGATCGTCTGAGCCTGTTAAACTTCATCT-----	304
Db	344	ThrLysGluLysValGluValProGluValIleValThrGlyAlaThrSerGluIleAla	363
QY	305	-----TCTTTGGCCACTGTGAAAAGCAACTGAGT	334
Db	364	GlnGlyThrProValCysAsnAspGlnProAsnLeuProGlnSerSerAspGlyIleArg	383
QY	335	AGATTCANATCAAGAAANACAGTGGAGTTACTCTGAACAAAGAGAGATTCGAACGGATC	394
Db	384	GlyValGluAsnSerLysThrAlaGluMetPro-----ArgHisThrIleLysLys---	400
QY	395	CACAGAGAAATAGCAATTCATAAAGCGCACAGTCTCTCCAATGGGACCTGCTGCTCT	454
Db	401	-----HisGluLysLysLysThrLysLysSerLeuMetLysArgLysLeuMet	416
QY	455	GAAGAACCGGACGAGCAGCAGTGGTTTTCCTCTGGAGAAAGGAGCCAGCATTC	514
Db	417	LysGluMetAlaGluLysProAlaAlaLeuLysGlnAlaGlu-GluGluAlaAlaLeuLys	436
QY	515	TCCATTCACATGTGCTCAGTGGCTGGAAG-----	545
Db	436	sglnAlaGluGluAlaAlaArgLysGlnAlaGluGluAlaAlaArgLysGlnAla	456
QY	546	-----GCAAGAACTCCCTGGAGCA	565
Db	456	agLluGluAlaAlaArgLysGlnAlaGluGluAlaAlaArgLysGlnAlaGluGln	476
QY	566	GGAAATTTTCACCTCTCCATAGAACAGCAGCAGCAGTGCAGACCCCTTTACTGACCCC	625
Db	476	uGluAla-----AlaArgLysGlnAlaGluGluAlaAlaArgLysGln	491
QY	626	TGTGAAAGGCTCTCTCCGAGCATGAGCTTACAGAGGCAAGATCGCAGCAGCAGA	685
Db	491	nAlaGluGluAlaAlaArgLysGlnAlaGluGluAlaAlaArgLysGlnAlaGln	511
QY	686	GCTTCAGAGGGCTCGGGCTCTGCAGCTCTACTATGANGCAGGCTCGAAGAGAGAAGA	745

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 11:21:37 ; Search time 12.2332 Seconds
(without alignments)
8801.668 Million cell updates/sec

Title: US-09-502-945-3
Perfect score: 2140
Sequence: 1 ggctgctgaatgactgcga.....gaagaagaagaagaaaaa 1298

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=framet_n2p_model -DEV=xlip
-Q/cgn2_1/USPTO.spool/US09502945/runat_14032003_101058_19100/app_query.fasta_1.10979
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40_cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945.qcgn_1.1.113_8runat_14032003_101058_19100 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGECURY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	307.5	14.4		1178	1	MN4_YEAST	P36044 saccharomyc
2	298	13.9		1898	1	TRHY_HUMAN	Q07283 homo sapien
3	289.5	13.5		1407	1	TRHY_RABBIT	P37709 oryctolagus
4	269	12.6		1240	1	YNJ1_YEAST	P53935 saccharomyc
5	264.5	12.4		1391	1	MST2_DROHY	Q08696 drosophila
6	262.5	12.3		877	1	INCE_CHICK	P53352 gallus gall
7	259	12.1		1549	1	TRHY_SHEEP	P22793 ovis aries
8	248.5	12.1		771	1	CALD_CHICK	P12957 gallus gall
9	7.5	12.0		678	1	GARP_PLAFF	P13816 plasmodium
10	5	12.0		1382	1	IF3A_HUMAN	Q4152 homo sapien
11	5	11.9		1085	1	YAF4_SCHPO	Q09863 schizosacch
12	4	11.9		899	1	YMJ3_YEAST	Q04500 saccharomyc
13		11.7		929	1	YDM6_SCHPO	P87137 schizosacch
14		11.6		848	1	NFM_MOUSE	P08553 mus musculu
15		11.6		1359	1	ATRX_CAPEL	Q9u7e0 caenorhabdi
16		11.6		1344	1	IF3A_MOUSE	P23116 mus musculu
17		11.6		484	1	SFRB_HUMAN	Q05519 homo sapien
18		11.5		793	1	CALD_HUMAN	Q05682 homo sapien

19	244.5	11.4	845	1 NFM_RAT	P12839 rattus norv
20	243.5	11.4	1403	1 YDF3 SCHPO	Q10475 schizosacch
21	239	11.2	1020	1 NFM_HUMAN	P12036 homo sapien
22	235	11.0	705	1 TRDN_RABIT	Q28870 oryctolagus
23	234.5	11.0	488	1 CYL2_BOVIN	Q28092 bos taurus
24	234	10.9	1394	1 CNG4_BOVIN	Q28181 bos taurus
25	232	10.8	915	1 NFM_HUMAN	P07197 homo sapien
26	227.5	10.6	1002	1 IF2P_YEAST	P37730 saccharomyc
27	227	10.6	831	1 NFM_RAT	P16884 rattus norv
28	226.5	10.6	2492	1 ATRX_HUMAN	P46100 homo sapien
29	226	10.6	978	1 RA50_AQUAE	O67124 aquifex aeo
30	226	10.6	1220	1 IF2P_HUMAN	P60841 homo sapien
31	224.5	10.5	644	1 NFM_RABIT	P54938 oryctolagus
32	224	10.5	494	1 SFR4_HUMAN	Q08170 homo sapien
33	222.5	10.4	810	1 NFM_BOVIN	O77788 bos taurus
34	221.5	10.4	728	1 TRDN_HUMAN	Q13061 homo sapien
35	220.5	10.3	344	1 MST1_DROHY	O08695 drosophila
36	219.5	10.3	805	1 IF2_AQUAE	O67825 aquifex aeo
37	218	10.2	471	1 RUL7_XENLA	P09406 xenopus lae
38	214.5	10.0	700	1 TRDN_CANFA	P82179 canis famil
39	214.5	10.0	911	1 CAFA_MOUSE	Q9qwf0 mus musculu
40	213	10.0	829	1 TQ1_XENLA	P41512 xenopus lae
41	212	9.9	185	1 T2_MOUSE	O06666 mus musculu
42	212	9.9	348	1 CYL2_HUMAN	Q14093 homo sapien
43	212	9.9	448	1 RUL7_DROME	P17133 drosophila
44	211.5	9.9	767	1 TOPI_CRIGR	Q07050 cricetus
45	211.5	9.9	2468	1 MAPE_HUMAN	P46821 homo sapien

ALIGNMENTS

RESULT 1

MN4_YEAST
ID MN4_YEAST STANDARD; PRT; 1178 AA.
AC P36044; P36043; P89095;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 18-OCT-2001 (Rel. 40, Last annotation update)
DE MN4 protein.
GN MN4 OR YKL200C/YKL201C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=97175967; PubMed=9023541;
RA Odani T., Shimma Y.-I., Tanaka A., Jigami Y.;
RT "Cloning and analysis of the MN4 gene required for phosphorylation
of N-linked oligosaccharides in Saccharomyces cerevisiae.";
RL Glycobiology 6:805-810(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Maia e Silva A., Bossier P., Villela C., Fernandes L., Soares H.,
RL Guerreiro P., Rodrigues-Pousada C.;
CC Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: MAY FUNCTION AS A POSITIVE REGULATOR FOR
MANNOSYLPHOSPHATE TRANSFERASE. IS REQUIRED TO MEDIATE
MANNOSYLPHOSPHATE TRANSFER IN BOTH THE CORE AND OUTER CHAIN
PORTIONS OF N-LINKED OLIGOSACCHARIDES.
CC -! SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -! SIMILARITY: TO YEAST YJR061W.
CC -! CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

[illegible]

ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS. DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG DIFFERENT SPECIES.

-1- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.

-1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100 FAMILY.

-1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: L09190; AAA65582.1; -;
 PIR: A45973; A45973.
 HSP: P02633; 4ICB.
 Genew; HGNC:11791; THH.
 MIM; 190370; -;
 InterPro; IPR001751; CaBP_S100.
 InterPro; IPR002048; EF-hand.
 InterPro; IPR002017; Spectrin.
 Pfam; PF00036; efhand; 1.
 Pfam; PF01023; S_100; 1.
 ProDom; PD003407; CaBP_S100; 1.
 ProSITE; PS00018; EF_HAND; 1.
 ProSITE; PS00303; S100_CaBP; 1.
 Keratinization; Repeat; Calcium-binding.
 DOMAIN 1 91 S-100 LIKE.
 CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
 CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
 DOMAIN 314 390 6 X 13 AA TANDEM REPEATS OF R-R-E-Q-E-E-R-E-Q-Q-L.
 REPEAT 314 326 1-1 (APPROXIMATE).
 REPEAT 327 339 1-2 (APPROXIMATE).
 REPEAT 340 351 1-3 (APPROXIMATE).
 REPEAT 352 364 1-4.
 REPEAT 365 377 1-5.
 REPEAT 378 390 1-6.
 DOMAIN 391 444 9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
 REPEAT 391 396 2-1.
 REPEAT 397 402 2-2.
 REPEAT 403 408 2-3.
 REPEAT 409 414 2-4.
 REPEAT 415 420 2-5.
 REPEAT 421 426 2-6.
 REPEAT 427 432 2-7.
 REPEAT 433 438 2-8.
 REPEAT 439 444 2-9.
 DOMAIN 444 702 9 X 28 AA APPROXIMATE TANDEM REPEATS.
 REPEAT 923 1162 8 X 30 AA TANDEM REPEATS.
 REPEAT 923 952 4-1.
 REPEAT 953 982 4-2.
 REPEAT 983 1012 4-3.
 REPEAT 1013 1042 4-4.
 REPEAT 1043 1072 4-5.
 REPEAT 1073 1102 4-6.
 REPEAT 1103 1132 4-7.
 REPEAT 1133 1162 4-8.
 DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDEM REPEATS.
 REPEAT 1752 1794 F -> L (IN REF. 2).
 CONFLICT 1794 1801 QERDQIR -> RSETGSTG (IN REF. 2).
 CONFLICT 1857 1880 Q -> K (IN REF. 2).
 CONFLICT 1880 1880 V -> G (IN REF. 2).
 SEQUENCE 1898 AA; 247219 MW; A74B5947FB62E31D CRC64;

Alignment Scores:
 Pred. No.: 1.13e-11 Length: 1898
 Score: 298.00 Matches: 96
 Percent Similarity: 48.10% Conservative: 106
 Best Local Similarity: 22.86% Mismatches: 136
 Query Match: 13.93% Indels: 82
 DB: 12 Gaps: 12

US-09-502-945-3 (1-1298) x TRHY_HUMAN (1-1898)

QY	77	CCAAAGACACCTCTTGTGAGTGAAGATCAGGGGGGACATGATGAGAGAGAAG	136
DB	124	ProargaspargGlnLeuGluGlu-----GluProGlyGlnArgArg	137
QY	137	CATCNAAAGCTCTTGAAGCAATCAGTCCCTTGATGAAAGAAATAGCCGAATTCGGCT	196
DB	138	ArgGlnLysArgGlnGlu-----GlnGluArgGluLeuAla	149
QY	197	GANAGGTCTGAGCTAGTCTGAAGGTCTCAGAGTTCAATGTCAGTCTGAAGGATCAGGA	256
DB	150	gluglyglugluGlnSerGluLysGlnGluGluGlnArgAspArgGlnArgArg	169
QY	257	GAAGAGCTGCTCTTGCAGATCTGTGAGCCTGTTAAACATTCATCTTCTTGGCCACT	316
DB	170	AspGluLeuTrpArgGlnArgGlnGluTrpGlnGluArgGluGluArgArgAlaGlu	189
QY	317	GTGAAAGCACTGAGTAGAGTCANATCAAGAAACAGTGGAGTTACCT-----	367
DB	190	---GluGluGlnLeuGlnSerCysLysGlyHisGluThrGluGluPheProAspGluGlu	208
QY	368	---CTGACAAAGAAAGAGATTGACCGATCCACAGAGATAGCATTAATAAACGACA	424
DB	209	GlnLeuArgArgArgGluLeuGluGlnLeuArgArgLysGlyArgGluGluLysGln-Gl	228
QY	425	AGTCTCTCCAAATGGGACCTCTGCTCTGAGAACCGCGCAGCAGACGAGCTGCTTT	484
DB	228	n-----GlnArgArgGluArgGlnArgGlnArgValPh	238
QY	485	TCCCTCTGGAGAAAGAGAGAGCCATTCCTCCCATTAACATGCTCAGTGGCTGGA	544
DB	238	eGlnGluGluGluLysGlu-----ThrValLeuArgLysGluGluLysLe	260
QY	545	GGCAAGAACTCCCTCGAGCAGGAGAAATTTCAACCTCTCCATAAGAACAGCAGCAGT	604
DB	247	GlysArgGlu-----	260
QY	605	GACAGACCTTTACTGACCCCTGTGGAAGGCTCTCTCCGAGCCATGAGCTAGAACGA	664
DB	260	uGlnGluGlu-----GluProGlnArgGlnArgGlnGluGlnGluGlnGlnLe	278
QY	665	GGCAAGATCGCACGACAGAGCTTCAGAGGGCTCTCGAGCTCTACTATGANGC	724
DB	278	uArgLysLeuGluArgGlnGluLeuArgArgGluArgGlnGluGlnGlnGlnGlnGln	298
QY	725	CAAGGCTCGAAGAGAGAGAAATCAAAAGTTAAAGTATCAAAAGTCGTAAGAAAGG	784
DB	298	nArgLeuArgGluGlnGlnLeuArgArg-----	308
QY	785	AAAGGCCAAGAAAGCCCTAAAGAGTTTTCAGCAGCTGCGGAGGTTAATCCAGCTCCGC	844
DB	309	-----LysGlnGluGluArgArgGlu-----	316
QY	845	ACTAGAAGAACGAAGAA	904
DB	317	-GlnGlnGluGluArgArgGluGlnGlnGlnGluArgArg-GluGlnGlnGluGluArgArg	336
QY	905	GAAGAAGAAAG	964
DB	336	luGlnGlnLeuArgArgGluGlnGluGluArgArgGluGlnGluGlnGlnGlnGlnGln	356
QY	965	AGAAAAG	1021
DB	356	luGluGluArgArgGluGlnGlnLeuArgArgGluGlnGluGluGluGlnGlnGlnGln	376


```

QY 620 GACCCTGTGGAAGAGCTCTCTC-----CGAGCCATGAGCGCTAGA 661
D 919 uGlnGluSerGluGluGluArgLeuArgGlnGluArgGluArgGluArgGluGlu 939
QY 662 AGAGCAAGATGCGACGA-----GCAGAGCTTCAGAGGCTCGGGCTCTGCATCCTA 715
D 939 uGluGlnLeuLeuArgGluGluGlnGluArgGluArgGluArgGluArgGlu 955
QY 716 CTATGANGCCAGGCTCCAGAGAGAGAAATCAATCAAGTTAAAGTAGTATCAAAAGTCGT 775
D 956 -----ArgLysLeuArgGluGluGlnGluLeuGlnGluArgGluGluArgLe 973
QY 776 GAAGAAAGAAAGGCAAGAACCCCTAAAGAGTTTCAGCAGCTG----- 821
D 973 uArgArgGlnGluArgGluArgLysLeuArgGluGluGlnGluLeuLeuArgGluGlu 993
QY 822 -----CGAAGGTTAATCCAGCTGCGCCGACCTAGAGAACG 856
D 993 uGlnGluLeuArgGlnGluArgAspArgLysPheArgGluGluGlnGluLeuGlnGlu 1013
QY 857 AAGAAAGAGGAAAGAGGAGG----- 881
D 1013 uArgGluGluArgLeuArgArgGlnGluArgGlnGluArgGlnGluArgGlnGlu 1033
QY 882 -----AGAAGAGAGAGAAACAAGGAGAGAA 907
D 1033 gGlnLeuArgGlnGluLeuGluGluGlnPheArgGlnGluArgAspArgLysPheAr 1053
QY 908 GAAAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 963
D 1053 gLeuGluGluGlnGlnLeuArg-GlnGluLysGluGluLysGlnLeuArgArgGlnGluArg 1073
QY 964 -----AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1012
D 1073 sPaArgLysPheArgGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1093
QY 1013 GAAGAGAGAGAGAG----- 1026
D 1093 rGArgGluArgAspArgLysPheArgGluGluGluGlnGlnGlnGlnGluArgGluGlu 1113
QY 1027 --AGAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1084
D 1113 luArgLeuArgGlnGlnGluArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1133
QY 1085 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1138
D 1133 rGlnGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1153
QY 1139 GGAACNAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1192
D 1153 euGlnGluSerGlu--GluGluArgLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGln 1172
QY 1193 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1252
D 1172 luGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1192
QY 1253 GAAGAGAGAGAGAG-----TATAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
D 1192 rGlnGluArgGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1210

```

RESULT 4
YNJ1_YEAST STANDARD: PRT; 1240 AA.
AC PS3935;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 141.5 kDa protein in YPT53-RHO2 intergenic region.
GN YN091W OR N231.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

```

[1]
SEQUENCE FROM N.A.
STRAIN-S288c / FY1679;
MEDLINE-96367601; PubMed-8771715;
RA Garcia-Cantalejo J.M., Boskovic J., Jimenez A.;
RT "Sequence analysis of a 14.2 kb fragment of Saccharomyces cerevisiae
chromosome XIV that includes the ypt53, trnAlu and gsr m2 genes and
four new open reading frames.";
RL yeast 12:599-608(1996).
CC -!- SIMILARITY: TO S.POMBE SPAC29E6.10C.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; X85811; CAA59826.1; -
DR EMBL; Z71367; CAA95967.1; -
DR SGD; S0005035; YN091W.
KW Hypothetical protein.
FT DOMAIN 756 761 POLY-GLU.
SQ SEQUENCE 1240 AA; 141513 MW; 3FE9D265822D5778 CRC64;

Alignment Scores:
Pred. No.: 7,43e-10 Length: 1240
Score: 269.00 Matches: 107
Percent Similarity: 41.42% Conservative: 74
Best Local Similarity: 24.49% Mismatches: 152
Query Match: 12.57% Indels: 104
DB: 11 Gaps: 11

US-09-502-945-3 (1-1298) x YN01_YEAST (1-1240)
QY 26 CTGTCAGAGAGCTCTGCGCTTTGAGCCANCAGAGAACTAGCGGATTTCGCAAGAGAC 85
D 431 LeuGlyProSerValLeuAlaAlaThrLeuSerSerCysPheSerGlnSerLysAsp 450
QY 86 TACCTCTTGAGT-----GAGAGTGAAGATGAGGGGACAAATGATGGAGAG 130
D 451 ThrSerLeuAspThrAspSerIleTyrGluAspGluAspGluGluAspTyrAspTyr 470
QY 131 AGAAGCATCAAAGCTTCTGGAAGCAATCAGTTCCCTTGATGGAAGAAATAGCGGAAA 190
D 471 SerGluTyrAlaGluAspSerGluGluValSerGluTyrGluGluGluGlu----- 487
QY 191 TTGGCTCANAGGCTGAGGCTAGTCTGAAGGTCTCAGAGTTCATGTCAGTTCGAGGA 250
D 488 -----AlaValGluLysProGlnHisAspGluLysSerAsnGly 500
QY 251 TCAGGAGAAAAGCTGGTCTTCGAGATCTGCTTGAAGCTCTTAAACTTCA----- 301
D 501 IleArgGluThrLeuHisLeuSerTyrAspHisAspHisLysArgGlnHisProHis 520
QY 302 -----TCTTCTTTGGCCACTGTGAAAAGCACTAGTAGTAGTCANATCAAAG 349
D 521 HisHisTyrHisSerThrHisSerGluAspGluLeuSerGluGluGluTyrIle 540
QY 350 AANACAGTGAGTTACCTCTGACAAAGAGAGATGAACGGATCCACAGAGAAATAGCAT 409
D 541 SerAspIleGluLeuPro-----HisAspProHisLysHisPheHisArgAsp----- 556
QY 410 TCAATAAAGCACACAGTCTCTCCAAATGGACCCTGCTCTCTCTGAGAACCGCGAGC 469
D 556 ----- 556
QY 470 AGAGCAGCTGGTTTTTCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 529
D 556 ----- 556
QY 530 GTCAGTGGCTGGAAGGCAAGAACTCCCTCGAGAGAGAGAGAGAGAGAGAGAGAG 589

```



```

Db 557 AspAspIleLeuAspGlyAsp----- 563
QY 590 GAACAAGCAGCAGTACAGACCCCTTACTGACCCCTGTGGAAAGGCCCTCTCTCCGAGC 649
Db 564 -----GluAspGluProGluGluGluAspGluAsnGluGluGlyAspAspGluGlu 579
QY 650 CATGACCCTAGAAGAGCAAGATGCGAGCAGCAGAGCTTCAGAGGCTCGCGCTCGCA 709
Db 580 AspThrTyAspSerGlyLeuAspGluThrAspArgLeuGluGluGlyArgLysLeuL 599
QY 710 GTCCTACTATGCGCAAGCTCGAAGAGAGAGAAATCNAAGTATAAAGTATCACAA 769
Db 599 eGlnIleAlaIleThrLysLeuLeuGlnSerArgIleMetAlaSer-----TyrHisG 617
QY 770 AGTCGTGAAGAAAGGAAAGCCCAAGAAAGCCCTTAAAGAGTTTGACACAGCTCGGAGGT 829
Db 617 uLysGlnAlaAspAsnAsnArgLeuLysLeuGlnGluLeuGlu----- 632
QY 830 TAATCCAGCTGCCGCTACTAGAAAGCAAGAAAGAGAGAAAGAGGAGGAGAGAAAGAA 889
Db 633 -----GluGluLysArgLysLysLysLysLysLysLysLysLysLysLysLysLys 646
QY 890 GAAGACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 949
Db 646 nLysLysArgGluLysGluLysGluLysLysLysLysLysLysLysLysLysLysLys 666
QY 950 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1009
Db 666 uLys-ArgLys-ArgLys-ArgLys-ArgLys-ArgLys-ArgLys-ArgLys-ArgLys- 685
QY 1010 TAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1069
Db 685 LuMetArgArgGluAlaGlnArgLysLysValGluGluAlaLysArgLysLysAspG 705
QY 1070 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
Db 705 LuGluArgLysArgGluGluGluGlnGlnArgArgGluGluGluMetGlnGluLysGln 725
QY 1118 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1177
Db 725 rGlyGlnLysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 745
QY 1178 GAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
Db 745 lNysArgLeuGluGlnGluLysLeuGlnLysGluLysGluLysGluGluGluArgL 765
QY 1238 AAGTA---GAAGCGGAGAGAGAGAGAGAGAGAGATATAGAGAGAGAGAA 1281
Db 765 euIleAlaGluAspAlaLeuArgLysGlnLysLeuAsnGluGluGln 780

```

RESULT 5

MST2_DROHY

```

ID MST2_DROHY STANDARD; PRT; 1391 AA.
AC Q08696;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein mst101(2).
GN MST101(2).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=95045538; PubMed=7957199;
RA Neesen J., Padmanabhan S., Buenemann H.;
RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid
RT motif representing the major component of the sperm-tail-specific
RT axoneme-associated protein family Dhmst101 form extended
RT alpha-helical rods within the extremely elongated spermatozoa of

```

```

RT Drosophila hydei."
RL Eur. J. Biochem. 225:1089-1095(1994).
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
CC SPERMATIDS.
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
CC REPEATS.

```

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; X73481; CAA51876.1; .
DR PIR; S34154; S34154.
DR FlyBase; FBgn020733; Dhyd\mst101(2).
KW Sperm; Repeat; Multigene family; Polymorphism.
FT DOMAIN 332 1368
FT [KR]-K-X-C-X-X-K-X-K-X-K-X-X-E.
SQ SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48878 CRC64;

```

Alignment Scores:

```

Pred. No.: 1,4e-09 Length: 1391
Score: 264.50 Matches: 112
Percent Similarity: 42.86 Conservative: 71
Best Local Similarity: 26.23 Mismatches: 175
Query Match: 12.36 Indels: 69
DB: 1 Gaps: 11

```

US-09-502-945-3 (1-1298) x MST2_DROHY (1-1391)

```

QY 128 GAGAGAAAGCATCNAAGCTTCTGGAAGCA-----ATCAGTTCCTCTGATGAAAG 178
Db 265 LysArgValIleAlaGluMetCysGluAlaAlaGlyTyrvLysSerSerGluProLys 284
QY 179 AATAGCGGGAATTTGGCTGANAGTCTGAGGCTAGTCTGGAAGGTCTCAGTTCATGTC 238
Db 285 LysLysGlyLysLysLysLysAsnAspGluLysLysGluLysGluLeuGluArgGlu 304
QY 239 AGTTCCTGAGGATCAGGAGAAAGCTGGTCTGCAGATCTCTTGAGCCTCTTAAA--- 295
Db 305 LeuLysGluGlnAlaGluGluGluAlaLysIleArgGlyValValLysLysLys 324
QY 295 ----- 295
Db 325 LysCysLysGluLysAlaLeuLysLysLysCysLysAspLeuLysMetLysGlu 344
QY 296 -----ACTTCATCTCTTTGGCCACTGTGAAAGCAACTGACTAGACTCANATCA 346
Db 345 GluAlaGluLysLysLysCysAlaAlaLeuAlaLysLysGlnLysGluGluAspGlu 364
QY 347 AAGAANACAGTGGAGTTACCTCTGAACAAGAAAGAGATT---GAACGGATCCACAGAGA 403
Db 365 LysAlaCysLysGluLeuAlaLysLysLysLysLysGluAlaAspGluLysLysCys 384
QY 404 TAGCATTTCAATAAAGCGCAACAGTCTCTCCAAATGGGACCTGCTGCTCTGAAGAAC 463
Db 385 GluAlaAlaAsnLysGluLysLysAlaAlaGluLysLysLysLysLysAlaAla-Ly 404
QY 464 GCAGCAGACAGCTGGTGTTCCTCCCTGGAGAAAGAGAGAGCCAGCCATTCCTCCCAT 523
Db 404 sGluArgLysGluAlaAlaGluLysLysLysCysGluGluAlaAla----- 419
QY 524 ACATGTGCTCAGCTGGCTGGAAGGCAAGAACTCCCTGGAGCAGGAAATTTTCAACCT 583
Db 420 -----LysLysGluLysGluAlaAlaGluArgLysLysCysGluGlu 434

```


Db	366	ProProValProGlnThrProProLysLeuAspPheGlnGlyLeuArgMetSerLeuArg	385
QY	308	TTGGCCACCTGTGAAAAAGCAACTGAGTAGTCANATCAAAAGAAANACAGTGGAGTTACCT	367
Db	386	SerGlnThrValAsnArgAsnGluGlnGlnGlnThrSerAsnAsn---GluCysAsp	404
QY	368	CTGAACAAA-----GAAGAGATTGAACGGATCCACAGAGAAATAGCATTTCAATA	415
Db	405	LeuSerLysSerGlnLysThrGlnGluProGlnSerAlaArgArgLysThrSerTyr	424
QY	416	AAACGGCAAGTCCTCTCCAAATGGGACCTGTGCTCTGAAGAACCGGCAGACGACGACA	475
Db	425	LysArgAla-ValAspGlnArgTyrAsp-----ThrGlnGlnAlaGluAs	439
QY	476	G---CTGGTTTTCCTCCGACGAAGAGAGCCAGCC-----	509
Db	439	pGlyGlyLeuSerProLeuArgLysLysThrProSerProCysProAlaSerLysVa	459
QY	510	-ATTGCTCCCATTTGAACATGTGCTC-----	533
Db	459	lValArgProPheLysThrPheLeuHisThrValGluLysAsnGlnLeuLeuMetThrPr	479
QY	534	-----AGTGGCTGGAAGCAGCAAGACTCCCTGGGA	562
Db	479	oSerSerValIlyArgAsnGlyValIleLysSerPheIleLysThrAsnThrProLeuG	499
QY	563	GCAGGAAATTTTCAACCTCTCCATAAGAACAGCAGCCAGCTGACAGACCCCTTTACTGAC	622
Db	499	nHisAsp-----ProLysGlnLysGluArgGlnLysLeuGlnAlaLeuAr	514
QY	623	CCCTGTGGAAAGGCC---TCTCTCGGACCATGAGCCTAGAAGAGCAAGATCGCAGC	679
Db	514	gLysLysGluGluAlaGluGlnLeuArgLysGlnLysValGluGluGluLysArgArg	534
QY	680	AGCAGAGCTTCAGAGGGCTCGGGCTCTGCAGTCTTACTATGANGCCAAAGGCTCGAAGAGA	739
Db	534	gGlnGlu-----GluAlaLysLeuArgArgG	543
QY	740	GAGAAATCNAAAGTTAAAGTATCACAAAGTCTGAAGAAAGGAAGGCCAAGAACG	799
Db	543	uGluArgLeu-----ArgLysValLeuGlnAlaArgGluArgAlaGluG	558
QY	800	CCTAAAGAGATTTGACAGCGTCGGGAAGGTTAATCCAGCTGCGCCACTAGAGAAGCAAG	859
Db	558	nLeuGluGlu--GluArgLysArgArgIleGluGlnLysLeuAlaLeuPheAspGluLy	577
QY	860	AAAGAGGAAAGAAAGAGGAGGAGA-----AGAAGAAGAAACAAGGA	901
Db	577	sThrGluLysAlaArgGluGluArgLeuAlaGluGluLysIleLysLysArgAlaAla	597
QY	902	GAAGAAGAAAGAA-----CAAGGAGAGAGAGAGAAAGAGGAGAGAGAAAGGA	955
Db	597	aLysLysMetGluGluAlaGluAlaArgArgArgGlnAspGlu-GluAlaArgLysGlnL	617
QY	956	AGAGAGGAAAGAAAGAGAGAGAAAGAAAGAAAGAGAG-----AAGAAG	1003
Db	617	ysAlaLeuGlnGlnGluGluGluArgArgHisLysGluLeuMetGlnLysLysLysG	637
QY	1004	AAGNACTAAGAAGAGAGAGGAGAGAAATAGAAAGAAAGAGAAAGAAAGTNAAGA	1063
Db	637	luGluGluGlnGluArgAlaArgLysIleAlaGluGlnArgGlnAlaGluGlnGluArgG	657
QY	1064	AGAAAGAAAGAGGAAAGAGAGAGAGAAAGTNAAGAGAGAAAGAGAGAGAAAG	1123
Db	657	luLysGlnLeuAlaAlaGluArgGluGlnGluArgLysLysGlnGluArgLysLysG	677
QY	1124	AAGAAAGAAAGAAATAGGAACNAGAAAGAGAGAGAAAGAAATAGAAAGAGAGAAAGAA	1183
Db	677	luGluGluArgIleGlnAlaGluLysGlnArgGluGlnGlnGlnLysAlaAlaArgLeuG	697
QY	1184	AAAGAGA-----AAAGAGAGAGAGAAAGAG-----GAGAAAAAGAGAGAAA	1225

Db 697 lnyGluValLeuAlaAlaLysGluGlnLysGluMetGlnLysLysGluLysG 717

Qy 1226 AAAGGAGAGAAAGTAGAAGCGGAAGAAAGAAAGAACTATTAAGAAAGGAAG 1285

Db 717 lnyGluGlnLeuAlaGluMetLysArgGlnGluGlnLysLysLeuProGluG 737

Qy 1286 AAGAAGG 1293

Db 737 lnyGlnLys 739

RESULT 7

TRHY_SHEEP STANDARD; PRT; 1549 AA.

AC P22793;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Trichohyalin.

GN THH.

OS Ovis aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;

OC Bovidae; Caprinae; Ovis.

OX NCBI_TaxID:9940;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93260018; PubMed=7680401;

RA Fietz M.J., McLaughlan C.J., Campbell M.T., Rogers G.E.;

RT "Analysis of the sheep trichohyalin gene: potential structural and

RT calcium-binding roles of trichohyalin in the hair follicle.";

RL J. Cell Biol. 121:855-863(1993).

[2]

RN SEQUENCE OF 1016-1549 FROM N.A.

RP STRAIN=Merino-Dorset horn X Border Leicester; TISSUE=Wool follicles;

RC STRAIN=Merino-Dorset horn N.A.

RX MEDLINE=90130632; PubMed=2298812;

RA Fietz M.J., Presland R.B., Rogers G.E.;

RT "The cDNA-derived amino acid sequence for trichohyalin, a

RT differentiation marker in the hair follicle, contains a 23 amino acid

RT repeat.";

RL J. Cell Biol. 110:427-436(1990).

CC -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES

CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE

CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR

CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY

CC ISOPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER

CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN

CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN

CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL

CC DIFFERENTIATION.

CC -!- SUBUNIT: HOMODIMER (PROBABLE).

CC -!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS MAY BE PRODUCED BY

CC ALTERNATIVE SPLICING OF THE SAME GENE.

CC -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS

CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN

CC THE EPITHELIA OF THE TONGUE, HOOF AND RUEN.

CC -!- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND

CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST

CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS

CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED

CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS

CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.

CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN

CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG

CC DIFFERENT SPECIES.

CC -!- PM: SUBSTRATE OF TRANSGLUUTAMINASE. SOME 200 ARGININES ARE

CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.

CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100

CC FAMILY.

CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collabora-

CC between the Swiss Institute of Bioinformatics and the EMBL outstati-

CC the European Bioinformatics Institute. There are no restrictions on

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL; 218361; CAA79165.1; -
CC EMBL; X51695; CAA35992.1; -
CC PIR; A34209; A34209.
CC PIR; S32633; S32633.
CC PIR; A40691; A40691.
CC HSSP; P02633; 1IG5.
DR InterPro; IPR001751; CaBP_S100.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 1.
DR Pfam; PF01023; S_100; 1.
DR ProDom; PD003407; CaBP_S100; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00303; S100_CaBP; FALSE_NEG.
KW Keratinization; Repeat; Calcium-binding.
FT DOMAIN 1 91
FT CA_BIND 22 33
FT CA_BIND 62 73
FT DOMAIN 413 832
FT REPEAT 413 448
FT REPEAT 449 476
FT REPEAT 477 504
FT REPEAT 505 532
FT REPEAT 533 560
FT REPEAT 561 588
FT REPEAT 589 616
FT REPEAT 617 644
FT REPEAT 645 678
FT REPEAT 679 706
FT REPEAT 707 742
FT REPEAT 743 771
FT REPEAT 772 796
FT REPEAT 797 832
FT REPEAT 833 860
FT DOMAIN 938 1507
FT REPEAT 938 961
FT REPEAT 962 985
FT REPEAT 986 1021
FT REPEAT 1022 1044
FT REPEAT 1045 1067
FT REPEAT 1068 1090
FT REPEAT 1091 1121
FT REPEAT 1122 1144
FT REPEAT 1145 1167
FT REPEAT 1168 1197
FT REPEAT 1198 1227
FT REPEAT 1228 1250
FT REPEAT 1251 1273
FT REPEAT 1274 1296
FT REPEAT 1297 1319
FT REPEAT 1320 1342
FT REPEAT 1343 1368
FT REPEAT 1369 1391
FT REPEAT 1392 1416
FT REPEAT 1417 1439
FT REPEAT 1440 1461
FT REPEAT 1462 1484
FT REPEAT 1485 1507
FT VARIANT 1145 1197
FT VARIANT 1251 1273
FT CONFLICT 1399 1399
SQ SEQUENCE 1549 AA; 201173 MW; E72FB9FF1326E54E CRC64;

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

Alignment Scores:

Pred No.: 3.05e-09
Score: 259.00
Percent Similarity: 43.81%
Best Local Similarity: 24.76%
Query Match: 12.10%
DB: 1

Length: 1549
Matches: 130
Conservative: 100
Mismatch: 161
Indels: 134
Gaps: 24

```

US-09-502-945-3 (1-1298) x TRHY_SHEEP (1-1549)
QY 47 TTGAGCCANAGAGAAAGTACGAGGATTTGCCAAAGACTACCTCTTGAGTGAGAGTAA 106
Db 716 LeuGlnArgGlnGluGluArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 735
QY 107 GATGAGGGGACAAATGATGAGAGAGAGAAAGCATCNAAG----- 145
Db 736 GluLysArgArgGlnValArgGluArgLysTyrLeuGluGluGlnGlnGluGlu 755
QY 146 ---CTTCTGGAAGCAATCAGTCCCTTGATGGAAGAAATAGCGGAAATTTGGCTGANAGG 202
Db 756 AspArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 775
QY 203 TCTGAGGCTAGTCTGAAGGTGTCAGAGTTCAATGTCAGTTCTGAA-----GGATCA 253
Db 776 GluLysValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 795
QY 254 GGAGAAAGCTGGTCTTCGAGATCTGCTTGAGCCTGTTAAACATTCATCTTTTGGCC 313
Db 796 ArgGluArgGlnTyrArgGluGlu-----GluLeuLeuArgGluGluArgLeuHis 813
QY 314 ACTGTGAAAAGCAACTGACTAGA-----GTCANATCAAGAANAACAGTGGAG 361
Db 814 ArgLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 833
QY 362 TTACCTCTGAACAAAGAGATTGAACGATC----- 394
Db 834 ArgGlnLeuGluGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 853
QY 395 -----CACAGAAATAGCATTCATAAACGACAGTCTCTCCAAA----- 437
Db 854 AspGlnHis-GlnAsnGluValArg---AsnSerArgValTyrSerLysHisArgGluAs 872
QY 438 -----TGGGACCTCTGCTCTGAAAGACCGGACGACGACGACGCT 478
Db 872 nLysGluLysSerArgGlnLeuAspSerTyrPValArgGluSerGlnPheGlnGlnAs 892
QY 479 GGTTCCTCCCTGGAGAAAGAGGAGCCATTCCTCCCATTCGACATGTGCTAGTGG 538
Db 892 pLeuArgProLeuGlnAspGluGlnGluGluLysArgGluArgGlu-----GlnG 909
QY 539 CTGAAGGCAAGAACT-----CCCTGGAGCAGGAAATTTTCAACCT 580
Db 909 uTyrArgSerArgGlnLysArgAspSerGlnPheProAlaGluGln-----Le 925
QY 581 CCTCCATAAGCAACAGCAGCAGTGACA-----GA 610
Db 925 uLeuGluArgGluGlnGlnGlnGlnGlnGlnGlnGluArgArgAspArgLysPheArgGluGlu 945
QY 611 CCCTTTACTGACCCCTGTGGAAAAGCCCTCTCCGAGCCATGAGCTAGAGAGCAAA 670
Db 945 uGlnLeuLeuLysGlyGlnArgGluGluLysIleArg-----TyrLeuGluGluAsp 963
QY 671 GATCCGACGAGCAGAGCTTCAGAGGGCTCGGGCTCTGAGTCTCTACTATGANGCCAAAGC 730
Db 963 gLysPheArgGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 983
QY 731 TCGAAGAGAGAGAAATCAAAAGTTAAAGATATCACAAGTCTGTAAGAAAGAAAGGC 790
Db 983 nGluArgAspArg-----LysPheArgGluGluLeuSerArgGlnGluArg 998
QY 791 CAAGAAAGCCCTAAAGAGTTTGAAGAGTTCGAGCAGCTGCGGAAGTTAATCCAGCTGCCACATAG 850
Db 998 gAspArgLysPheArgGluGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1010
QY 851 AGAACGAAGAAAGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 902
Db 1010 nGluArgGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1030
QY 903 -----AAGAAGAAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 943
```



```
|||||
245 sGluGluAla-----GluLysGluArgGluLysLeuGluAlaGluLys 260
QY TCAGAGGCTCGGCTCAGTCTACTATGANGC-----AAGCTCGAAG 736
Db sGluArgLeuLysAlaGluGluLysLysAlaAlaGluGluLysAlaGluGlu 280
QY AGAGAAGAAATNAAGTTAAAGTATCAAAAGTCGTGAAGAAAGAAAGCCCAAGAA 796
Db uGluLysLysAlaAla-----GluGluArgGluArgAlaLysAlaGluGluLys 298
QY AGCCCTAAAGAGTTTACAGCAGCTCGGAGGTATATCAGCTCCGCGACTAGAACG 856
Db gAlaAlaGluGluArgGluArg-----AlaLysAlaGluGluArg 312
QY AAGAAAGAGGAAAGAGAGGAGGAGAGAAAGAAAGAAAGAAAGAAAGAA 916
Db gLysAlaAlaGluGluArgGluArgAlaLysAlaGluGluArgLysAlaGluGlu 332
QY AGGAGAGAGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 973
Db ArgAlaLysAlaGluGluArgGluArgLysAlaAlaGluGluArgAlaLysAl 349
QY AAGAGAAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1033
Db aGluGluGluArgLysAlaAlaGluGluArgAlaLysAlaGluGluArgLysAla 369
QY GAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1090
Db aGluGluArgGluArgAlaLysAlaGluGluLysArgAlaAlaGluGluLysAla 389
QY AGGAAGAACTNAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1150
Db gLeuGluAlaGluLysLysLysLysLysLysLysLysLysLysLysLysLys 409
QY AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
Db uLysAlaGluAlaAsnLeuLeuArgLysGluGluGluGluGluGluGluGlu 429
QY GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1246
Db lalLysLysGluSerLeuProGluLysLysLysLysLysLysLysLysLysLys 449
QY GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1281
Db spAsnLysAspLysGluLysAlaProLysGluGlu 460

RESULT 9
GARP_PLAFF
ID GARP_PLAFF STANDARD; PRT; 678 AA.
AC PI3816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutamic acid-rich protein precursor.
GN GARP.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89040048; PubMed=2903445;
RA Trigila T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.;
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
RT acid-rich protein (GARP).";
RL Mol. Biochem. Parasitol. 31:199-202(1988).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J03998; AAA29605.1; -.
PIR: A54514; A54514.
KW Repeat; Malaria; Antigen; Signal.
FT SIGNAL 1
FT CHAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.
FT DOMAIN 120 164 15 X 3 AA TANDEM REPEATS OF K-K-X.
FT DOMAIN 372 416 9 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 417 441 5 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 576 604 POLY-GLU.
FT DOMAIN 605 653 7 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 654 663 POLY-GLU.
SQ SEQUENCE 678 AA; 80551 MW; 2A8F85608496EA9E CRC64;

Alignment Scores:
Pred. No.: 4.06e-09 Length: 678
Score: 257.50 Matches: 105
Percent Similarity: 39.91% Conservative: 77
Best Local Similarity: 23.03% Mismatches: 157
Query Match: 12.03% Indels: 117
DB: 1 Gaps: 14

US-09-502-945-3 (1-1298) x GARP_PLAFF (1-678)
QY 89 CTCCTGAGTGAAGATGAGGGGACATGATGAGAGAGAGAGAGAGAGAGCTT 148
Db 49 LeuLeuAsnGluThrGluLeuGluLysAsnLysAspAsnSerLysSerGluThrLeu 68
QY 149 CTG-----CAAGCAATCATCTCCCTGTGATGGAAGAGAGAGAGAGAG 187
Db 69 LeuLysGluLysAspGluLysAspValProThrThrSerAsnAspAsnLeuLys 88
QY 188 AATTTGGCTGANAGTCTGAGCTAGTCTGAAGGTGTGAGAGTTCATATGCTCTG 247
Db 89 AsnAlaHisAsnAsnGluLeuSerSerSerThrAspProThrAsnLeuLeuVal 108
QY 248 GGATCAGGAGAAAGCTGCTCTGAGATCTGCTGAGCTGTGAGCTGTAAACTCTCT 307
Db 109 AsnAspLysAspAsnGluAsnSerValAspLysLysLysLysLysLysLysHis 128
QY 308 TTGGCCACTGTCAAAAGCAACTGAGTGTAGAGTCANATCAAGAAACAGCTGGAT 367
Db 129 LysLysAspLysLysLysLysLysLysLysLysLysLysLysLysLysLys 148
QY 368 CTGAACAAGAAGAGATTGAACGGATCCACAGAGATAGCATTCAATAAAGCAGCA 427
Db 149 GluLysLysHisLysLysLysLysLysLysLysLysLysLysLysLysLys 168
QY 428 CCTCTC-----CAAATGGGACCTGCTGCTGTGAAGAACCCGCGAGGAGCA 475
Db 169 GluValMetSerLeuTyrLysThrGlyGlnHisLysProLysAsnAlaThr 185
QY 476 GCTGTTTTTCCCTGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
Db 186 -----GluHisGlyGluGluAsnLeuAspGlu-----Glu 195
QY 536 TGGCTGAAGAGCAAGAACTCCCTGGAGCAGGAGAAATTTTCAACCTCCTCCATA 595
Db 196 MetValSerGluLeuAsnAsnAlaGlnGlyLeu-----Leu 209
QY 596 GCAGCCAGTGCAGAGACCTTTACTGACCCCTGTGGAAGAGGC---CTCTCTCCGAG 652
Db 210 LeuSerSerProTyrGlnTyrArgGluGlnGlyCysGlyLeuLeuSerValHis 229
QY 653 GAGCCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 712
Db 230 Glu-----
QY 713 CTACTATGANGCAAGGCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 772
Db 231 -----ThrSerAsnAspThrLysAspAsnAspLysGluAsnLeuSerGluasp 246
```



```
QY 299 TCATCTTCTTGGCCACTGTGAAAGCAACAGTGAAGTGAAGTGCANATCAAGAAAGCAAGTG 358
Db 634 Lys-----ThrValArgGluArgLeuGluGlnle-----643
QY 359 GAGTTACCTCTGAACAAAGAGATGTGAACGATCCACAGAAATAGCATTCATTAATAA 418
Db 644 -----LysLysThrGlu-LeuGlyAlaLysAlaPheLysAspIle-----As 657
QY 419 CGCACAACTCTCCAAATGGACCTGTGTCCTGGAAGAACCGCAGCAGCAGCAAGCT 478
Db 657 pIleGluAspLeuGluLeuAspPheIleMetAlaLysGlnValGluGln-- 676
QY 479 GGTTCCTCCCTGGAGAAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 538
Db 677 -----LeuGluLysGluLysGluLeu-----684
QY 539 CTGAAGGCAAGCACTCCCTGAGCAGGAGAAATTTCAACCTCCTCCATAAGCAAGCA 598
Db 685 -----GlnGluArgLeuLysAsnGlnGluLysLysIleAspTyrPheGluArgAlaLys 703
QY 599 GCCAGTCACAGACCTTTACTG---ACCCCTGTGGAAAGGCCCTCTCTCCAGCCATGAG 655
Db 703 gLeuGluGluIleProLeuIleLysSerAlaTyrGluGluGlnArgIleLysAspMet 723
QY 656 CCTA-----GAAGAGCAAGATGCGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 703
Db 723 pLeuTrpGluGlnGlnGluGluArgIleThrThrMetGlnLeuGluArgGluLysAl 743
QY 704 TCTGCAGTCCTACTATGANGCCAGGCTCGAAGAGAGCAAGAAATCAATCAAGTAAAGTA 763
Db 743 aLeuGluHisLysAsnArgMetSerArgMetLeuGluAspArgAspLeuPheValMet 763
QY 764 TCACAAATCGTGAAGAAAGCAAGGCAAGCAAGCCTCAAGAGTTTTCAGCAGCTGCG 823
Db 763 gLeuLysAlaArgGlnSerValTyrGluGluLysLeuLysGlnPheGluArgLe 783
QY 824 GAAGTTAATCCAGCTCCCTGCTAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 883
Db 783 uAlaGluGluArgHisAsnArgLeuGluGluArgLysArgGluArgLysGluGluArg 803
QY 884 AAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 943
Db 803 gIleThrTyrTyrArgGluLysGluGluGluGluGluGluGluGluGluGluGluGlu 823
QY 944 AGAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1003
Db 823 uLysGluArgGluGluArgGluArg-----AlaGlu 833
QY 1004 AAGAACTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1063
Db 833 uArgAlaLysArgGluGluGluLeuArgGluTyrGlnGluArgValLysLysLeuGluGlu 853
QY 1064 AAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1123
Db 853 uValGluArgLysLysArgGlnArg-GluLeuGluIleGluGluArgGluArgArg 873
QY 1124 AAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1183
Db 873 luGluGluArgArgLeuGluLysSerSerLeuSerArgLysAspSerArgTrpGlyAsp 893
QY 1135 -----ATAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1176
Db 893 rgAspSerGluGluThrTrpArgLysGlyProGluAlaAspSerGluTrpArgArgGly 913
QY 1177 -----AGAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1219
Db 913 roProGluLysGluTrpArgArgGlyGluGluArgGluArgGluAspArgSerHisArg 933
QY 1220 AAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1279
Db 933 spGluGluArgProArgArgLeuGluLysAspGluAspArgGluProSerLeuArgPro 953
```

```
QY 1280 AAGAAGAAGA 1290
Db 953 spAspAspArg 956
RESULT 11
YAF_A_SCHPO STANDARD; PRT; 1085 AA.
ID YAF_A_SCHPO
AC Q09863;
DC 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C29E6.10c in chromosome I.
GN SPAC29E6.10c OR SPAC30.14c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabi C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Ruvellu J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -!- SIMILARITY: TO YEAST YNL091W.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 266525; CAA91432.1; -.
CC EMBL; AL136538; CAB66473.1; -.
KW Hypothetical protein.
FT DOMAIN 33 46 POLY-LYS.
FT DOMAIN 184 194 POLY-LYS.
FT DOMAIN 695 698 POLY-GLU.
FT DOMAIN 714 717 POLY-GLU.
FT DOMAIN 718 721 POLY-ARG.
FT DOMAIN 938 941 POLY-SER.
SQ SEQUENCE 1085 AA; 122940 MW; 041164132676F233 CRC64;
Alignment Scores:
Pred. No.: 5,19e-09 Length: 1085
Score: 255.50 Matches: 104
Percent Similarity: 42.40% Conservativity: 94
```


Best Local Similarity: 22.27% Mismatches: 158
 Query Match: 11.94% Indels: 111
 DB: 1 Gaps: 13

US-09-502-945-3 (1-1298) x YAFU_SCHPO (1-1085)

```

QY 59 GAAGAACTAGGAGATTTCGCCAAAGACTACCTCTTGAGTGAGAGTGAAGTGGGGGAC 118
Db 325 GluGlnThrAlaAsnIleGlnArgAsnLeuAlaAsnThrGluSerValAsnAlaSerAsp 344
QY 119 AATGATGAGAGAGAAAGCATCNAAGCTTCTGGAAGCAATCAGTTCCCTTGATGGAAG 178
Db 345 GluGlySerAsp-----Lys 349
QY 179 AATAGCGGAATATGGCTGANAGTCTGAGGCTAGTCTGAAGGTGTCAGAGTTCAATGTC 238
Db 350 SerGlnTyrGlyIleIleSerAspSerProLysLeuLeuSerIleProLeuAsnVal 369
QY 239 AGTTCTGAAGGATCAGGAGAAAGCTGCTCTTGCAGATCTGCTGAGCCCTGTTAAACT 298
Db 370 ProSerLysSerLeuAsnAspIleThrGlnAspGluLeu----- 383
QY 299 TCATCTCTTTGGCCACTGTGAAAAGCACTGAGTACAGTCAATCAAGAAACACAGTG 358
Db 384 AsnSerSerAsnAlaAspValAspGluValIleGluThrThrSerLeuGluGluLys 403
QY 359 GAGTTACTCTGACAAAGAGAGATTGAACGGATCCACAGAGATAGCATTTCAATAAAA 418
Db 404 AsnValAsp---AsnGlnGluPheValThrSerIleSerAsnGlyAsnGlnThrLeuGlu 422
QY 419 CGCACA-----AGTCCTCTCCAAATGGGACCTGCTCTCTGAGAACCGGAGGAGCA 472
Db 423 AspThrSerHisSerProGlnThrGlnProPheGlnProProTyrProSerLys--- 441
QY 473 GCAGCTGTTTTTCCCTCGGAGAAAGAGGAGCCAGCCATTGCTCCCAATTGAACATGCT 532
Db 441 ----- 441
QY 533 CAGTGGCTGGAAGGAGCAACTCCCTGGAGCAGAAATTTTCAACCTC----- 581
Db 442 -----AlaAspGluLysAsnSer-TyrHisSerAspLeuTyrAsnPheGlySerSerLe 459
QY 582 ----CTCCATAGAACAGCAGCCAGTGCACAGCCCTTACTGACCCCTGTGGAAGAGC 637
Db 459 uThrValLysGlyIleLeuThrValAlaAspLeuLeuLysAsnAspGlyLysLy 479
QY 638 CTCTCTCGAGCCATG---AGCCTAGAGAGGCAAGATCGAGCAGCAGAG--- 686
Db 479 sPheIleGluMetGluGlnLeuAlaGluArgMetGlnArgGluAspAsnSerAs 499
QY 686 ----- 686
Db 499 nPheHisGluProGluLeuTyrGluSerGlyLeuGluTyrAspGluAspGluGluAs 519
QY 687 -----CTTCAGAG 694
Db 519 pGluGluAspValAspGluAspGluLeuAspLeuMetThrAspGluGlnArgMetGluG 539
QY 695 GGCTCGGCTCTGAGTCTTACTATGANGCCAGGCTCGAAGAGAGAAATCNAAG 754
Db 539 uGlyArgArgMetPheGlnIlePheAlaAlaArgLeuPheGluGlnArgValLeuGlnAl 559
QY 755 TTAAGATATCAAAAGTCGTGAGAAAGAAAGCCCAAGAAAGCCCTAAAGAGTTTGA 814
Db 559 a-----TyrArgGluLysValAlaGlnGlnArgGlnAlaLysLeuLeuGluIleG 577
QY 815 GCAGCTGGGAGGTATATCCAGCTGCGCAGTACAGAACGACGAAAGAGAGAGAAAG 874
Db 577 uGluGluAsnLysArgLysGlnGluArgGluLeuLysIleArgGluLysGluLysLy 597
QY 875 GAGGAGGAGAAAGAGAGAAACAAGGAGAGAAAGAAAGAGAGAGAGAGAGAGAA 934
Db 597 sArgAspLysLysLysGlnLeuLysLeuAlaLysGluGluGlu-----ArgGlnArgAr 615

```

```

QY 935 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 994
Db 615 gGluAlaGluArgLeuAlaGluGlnAlaAlaGlnLysAlaLeu---GluAlaLysArgG 634
QY 995 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1052
Db 634 nGluGluAlaArgLysLysArgGluGluGlnArgLeuLysArgGluGlnGluLysLysG 654
QY 1053 -----AAAGTNAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1096
Db 654 nGlnGluLeuGluArgGlnLysArgGluGluLysGlnLysGlnLysGluArgGluLysLy 674
QY 1097 AACTNAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1156
Db 674 sLeuLysLysGlnGlnGlnGlnGluAlaAspArgGluLysMetAlaArgGluGlnArgLeuAr 694
QY 1157 AAGAAAGAAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1216
Db 694 gGluGluGluLysArgIleLeuGluGluArgLysArgGluLys----- 710
QY 1217 AGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1276
Db 711 -----LeuAspLysGluGluGluArgArg-ArgGluLeuLeuGluLysGluSerG 729
QY 1277 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1293
Db 729 LuGluLysGluArgArg 734

```

RESULT 12

```

YMJ3_YEAST
ID YMJ3_YEAST STANDARD; PRT; 899 AA.
AC Q04500;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 103.0 kDa protein in RAD10-PRS4 intergenic region.
GN YML093W.
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Gentles S., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC -1- SIMILARITY: TO C.ELEGANS C33G8.2 AND S.POMBE SPAC57A7.06.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z46660; CAA86645.1;
DR SGD; S0004558; YML093W.
KW Hypothetical protein; ATP-binding.
FT NP-BIND 260 267 ATP (POTENTIAL).
FT DOMAIN 162 168 POLY-SER.
FT DOMAIN 179 182 POLY-GLU.
FT DOMAIN 704 712 POLY-LYS.
SQ SEQUENCE 899 AA; 103023 MW; 0D4FC90D1CB3CFF1 CRC64;

```

Alignment Scores:

```

Pred. No.: 6.54e-09 Length: 899
Score: 254.00 Matches: 125
Percent Similarity: 41.03% Conservative: 67
Best Local Similarity: 26.71% Mismatches: 175
Query Match: 11.87% Indels: 101
DB: 1 Gaps: 22

```



```
US-09-502-945-3 (1-1298) x YMJ3_YEAST (1-899)
QY 95 AGTGAGAGTGAAGATGAGGGGACATGATGGAGAGAGAAAGCATCAAAAGCTTCGGAA 154
|||||
Db 171 SerGluSerGluSerGluSerGluSerGluSerGluSerGluSerGluSerGluSerGlu 190
QY 155 -----GCAATCAGTTCCTCCCTGATGAGAAAGATAGCGCG 187
Db 191 AspGluGluAspGluLeuLeuAsnThrIleThrSerLysLeuLeuAspGluThrLysSer 210
QY 188 AAATTGCTGCTGAGTCTGAG-----GCTAGTCTGAGGCTGCAGAGTTC----- 232
Db 211 LysAlaProLysArgLeuAspThrTyrglySerGlyGluAlaAsnGluTyryrValLeuPro 230
QY 233 -----AATGTCAGTTCCTGAGGATCAGGAGAAAGCTGCTGTCAGATCTGCTGAG 286
Db 231 SerAlaAsnAlaLaserGlyAlaSerGly--LysLeuSerLeuThrAspMetMetAsn 249
QY 287 CCTGTTAAACTTCATCTCTTTTG-----GCCACTGTGAAAAGCAACTGAGTAGATC 340
Db 250 ValIleAspAspArgGlnValIleGluAsnAlaAsnLeuLeuLysGlyLysSer----- 267
QY 341 ANATCAAGAAACAGTGGAGTTACCTCTGNACAAAGAGATGTAACGATCCAC--- 397
Db 268 -----SerThrTyrgluValProLeu---ProGlnArgIleGlnGlnArgHisAsp 283
QY 398 ---AGAAATAGCATTCATATAAACGACACAGTCTCTCCAAATGGGACCTGTCGCTCT 454
Db 284 ArgLysAlaAlaTyrgluIleSerArg-GlnGluValSerLysTrpAsnAspIleValG 303
QY 455 GAAGAACGGCAGCAGCAGCAGCTGTTTTCCTCCCTGGAGAAAGAGCAGCCACTGTC 514
Db 303 nGlnAsnArgArgAlaAspHisLeuIlePheProLeuAsnLys----- 317
QY 515 TCCCATTCAGCATGCTGCTCAGTCTCGAGGCGAAGA-----ACTCC 556
Db 318 -ProThrGluHisAsnHisAlaSerAlaPheThrArgThrGlnAspValProGlnThrG 337
QY 557 CCGGAGCAGGAGAAATTTCAACCTCCCTCCATAAGAACAAG-----CAGCCAGTGACAGA 610
Db 337 uLeuGlnGluLysValAspGlnValLeuGlnGluSerAsnLeuAlaAsnProGluLysAs 357
QY 611 CCCTTTACTGACCCCTGTGGAAAGGCTCTCTCCGAGCCATGAGCTGAGAGGAGCAAA 670
Db 357 p-----SerLysPheGluGluLeuSerThrAlaLysMetThrProGluGluMetAr 374
QY 671 GATCGCAGCAGCAGCTTCAGAGGCTCGGCTCTGCTCAGTCTCTACTACTGANGCCAGGC 730
Db 374 gLysArgThrThrGluMetArgLeuMetArgGluLeuMetPheArgGluArgLysAl 394
QY 731 TCGAAGAGAGAGAAATTCNAAAGTTAAAGTATACAAAGTCTGTAAGAAAGGAAAGGC 790
Db 394 aArgArgLeuLysLysLysLysSerLysThrTyrglyLysIle-----Ly 409
QY 791 CAAGAAGCCCTAAAGAGTTGAGCAGCTCGGAAGGTTAATCCAGCTGCCGCACTAGA 850
Db 409 sLysLysLysLysLysLysLysAsnArgGluLeuAlaValSerSerAspGluAspAsnG 429
QY 851 AGACGA-----AGAAAGAGGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 901
Db 429 uAspHisAspIleAlaArgAlaLysGluArgMetThrLeuLysHisLysThrAsnSerLy 449
QY 902 GAAGAAGAAAGAA-----GAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 952
Db 449 strAlaLysAspMetIleLysHisGlyMetThrAsnAspAlaGluThrArgGluGluMe 469
QY 953 GGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 1012
Db 469 tGluGluMetLeuArg-----GlnGlyGluArgLeuLysAlaLy 482
QY 1013 GAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 1066
|||||
-----AGAAAGAAAGAAAGTNAAGAAAGA 1066
|||||
```

```
Db 482 sMetLeuAspArgAsnSerAspGluCluAspGlyArgValGlnThrLeuSerAspVa 502
QY 1067 AGAAGAAGAGAAAGAGAA-----AGAAGAGAGAGAAAC 1099
|||||
Db 502 lGluAsnGluGluLysGluAsnIleAspSerGluAlaLeuLysSerLysLeuGlyLysTh 522
QY 1100 T-----NAGAAGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 1138
Db 522 rGlyValMetAsnMetAlaPheMetLysAsnGlyGluAlaArgGluArgGluAlaAsnLy 542
QY 1139 GGAACNAGAAAGAGGAGGAGAA-----AAGAATAAGAGAGAGGAGGAGGAGGAGGAA 1189
|||||
Db 542 sGluThrLeuArgGlnLeuArgAlaValGluAsnGlyAspAspIleLysLeuPheGluSe 562
QY 1190 GAAAGAAGAGGAGGAGGAGGAGAA-----AAAGGAGGAGGAGGAGGAGGAGGAGGAA 1226
Db 562 rAspGluGluGluThrAsnGlyGluAsnIleGlnIleAsnLysGlyArgIleTyrrTh 582
QY 1227 -----AAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 1258
Db 582 rProGlySerLeuGluSerAsnLysAspMetAsnGluLeuAsnAspHisThrArgLysGl 602
QY 1259 AAAGAAGAGTATAAGAGGAGGAGAA 1280
Db 602 uAsnLysValAspGluSerArg 609
RESULT 13
YDM6.SCHPO
ID YDM6.SCHPO STANDARD; PRT; 929 AA.
AC P87137;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C57A7.06 in chromosome I.
GN SPAC57A7.06
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Coffeau A., Cadieu E., Dreano S., Gloux S., Leleure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Batreil B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -!- SIMILARITY: TO YEAST YML093W.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



```
Db 612 lGluGluValLysProLysProGluAlaLysAlaGlyLysGlyGluGluLysGluGluGlu 632
QY 1121 AAGAAGAAAGAAATAGGACNAGA-AAGAAGGAGAAAGAAATAGGAGGAAGA 1179
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 632 uLysValGluGluGluLysGluValThrLysGluSerProLysGluGluValGlu 652
QY 1180 AAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1239
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 652 uLysGluGluLysProLysAspValAlaAspLysLysAlaGluSerProValLy 672
QY 1240 GTAGAAAGCGGAAGAAGA-----1258
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 672 sGluLysAlaValGluGluValIleThrIleSerLysSerValLysValSerLeuGluLy 692
QY 1259 -----AAGAAAGTATAGAAAGGAGAAAGAAAGAAAGAA 1290
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 692 sAspThrLysGluGluLysProGlnProGlnGluLysValLysGluLysAlaGluGlu 712
QY 1291 AGGA 1294
    ||||
Db 712 uGly 713
    ||||
RESULT 15
ID ATRX_CAEEL STANDARD; PRT; 1359 AA.
AC Q9U7E0; O02061;
DT 13-JUN-2002 (Rel. 41, Created)
DT 13-JUN-2002 (Rel. 41, Last sequence update)
DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN XNP-1 OR B0041.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID:6239;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99365296; PubMed=10433961;
RA Villard L., Fontes M., Ewbank J.J.;
RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RL the human XNP/ATR-X gene."
RL Gene 236:13-19(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton R., Woldmann P.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN (Potential).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF134186; AAD55361.1; -
DR EMBL; AF000196; AAC24256.1; -
DR WormPep; B0041.7; CEI7314.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00176; SNF2_N; 1.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.
FT NP_BIND 496 503
FT SITE 636 639
FT DEAH_BOX
```

```
FT DOMAIN 67 70 POLY-ASP.
FT DOMAIN 266 272 POLY-GLU.
FT DOMAIN 276 281 POLY-LYS.
FT DOMAIN 372 375 POLY-LYS.
FT DOMAIN 603 608 POLY-LYS.
FT DOMAIN 859 862 POLY-LYS.
FT CONFLICT 479 479 C -> F (IN REF. 2).
SQ SEQUENCE 1359 AA; 156191 MW; EB4342547D4F4E64 CRC64;

Alignment Scores:
Pred. No.: 1,29e-08 Length: 1359
Score: 249.00 Matches: 64
Percent Similarity: 52.43% Conservative: 44
Best Local Similarity: 31.07% Mismatches: 74
Query Match: 11.64% Indels: 24
DB: 1 Gaps: 5

US-09-502-945-3 (1-1298) x ATRX_CAEEL (1-1359)
QY 731 TCGAAGAGAGAAATCNAAGTTAAAGTATCAAAAGTCGTGAAGAAAGGAAGGC 790
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 29 AsnGluArgLysGluLysArgAlaGlnLysLeuLysGluLysArg---GluArgGluGly 47
QY 791 CAAGAAAGCCCTAAAGAGATTTCAGCAGCTGCCGAGGTTAATCCAGCTCCCGC-----844
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 48 LysProProLysLysArgProAlaLysLysArgLysAlaSerSerSerGluGluAsp 67
QY 845 -----ACTAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 871
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 68 AspAspGluGluGluSerProArgLysSerSerLysSerArgLysArgAlaLys 87
QY 872 AAGGAGGAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 931
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 88 SerGluSerGluSerAspGluSerAspGluGluAspArgLysSerLysSerLys 107
QY 932 AAAAGAGGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 991
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 108 Lys-----LysValAspGlnLysLysGluLysSerLysLysLysArg 122
QY 992 AGGAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1051
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 123 ThrThrSerSerGluAspGluAspSerAspGluGluArgGluGlnLysSerLys 142
QY 1052 AAAAGTNAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1111
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 143 LysSerLysThrLysLysGlnThrSerSerGluSerSerGluGluSerGluGlu 162
QY 1112 AGAGGAGGAAGAAAG-----AAAGAGAATAAGAACNAGAAAGAGGAGAAG 1159
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 163 ArgLysValLysSerLysLysAsnLysGluLysSerValLysLysArgAlaGluThr 182
QY 1160 AAAGAATAAGAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1219
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 183 SerGluLysSerAspGluAspGluLysProSerLysSerLysGlyLeuLysLys 202
QY 1220 AAGAAAAAGGAGAAAGAAAGTAAAGCGGAAAGAAAGAAAGAAAGTATAAGAGGAAG 1279
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 203 LysAlaLysSerGluSerGluSerGluSerGluAspGluLysGlu---ValLysLysSer 221
QY 1280 AAGAAAGAAAGAAAGAAAG 1297
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 222 LysLysLysSerLysLys 227
```

Search completed: March 21, 2003, 12:47:36
Job time : 34.4332 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:20:42 ; Search time 44.5135 Seconds
(without alignments)
12016.549 Million cell updates/sec

Title: US-09-502-945-3

Perfect score: 2140

Sequence: 1 ggcgtcgaatgactgcga.....gaagaagaagaaggaagaaaa 1298

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q-/cgn2_1/USPTO_spoil/US09502945/runat_14032003_101058_19113/app_query.fasta_1.10979
-DB-SPTREMBL_21 -QFMT-fastan -SUFFIX-rspt -MINMATCH=0.1 -LOEPCI=0 -LOEPTX=0
-UNITS=bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -List=45
-DOALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER-US09502945.ecgn_1.1.565.ernat_14032003_101058_19113 -NCPG=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organella:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.rvirus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1338.5	62.5	763	4 Q8WWS6	Q8WWS6 homo sapien

2	1338.5	62.5	771	4 Q9BVJ6	Q9BVJ6 homo sapien
3	1183	55.3	766	4 Q92555	Q92555 homo sapien
4	600	28.0	131	4 Q60531	Q60531 homo sapien
5	571	26.7	194	11 Q9CZK8	Q9CZK8 mus musculus
6	403	18.8	260	10 Q9IGZ9	Q9IGZ9 arabidopsis
7	365	17.1	554	5 Q9VPS3	Q9VPS3 drosophila
8	322.5	15.1	699	10 Q9ZU69	Q9ZU69 arabidopsis
9	312.5	14.6	1701	5 Q61164	Q61164 plasmodium
10	312	14.6	2274	5 Q9VVU0	Q9VVU0 drosophila
11	310.5	14.5	685	10 Q8W3X8	Q8W3X8 cucurbita m
12	309.5	14.5	771	5 Q9VUR2	Q9VUR2 drosophila
13	309.5	14.5	2055	5 Q8F5C7	Q8F5C7 plasmodium
14	302	14.1	2081	10 Q9LH98	Q9LH98 arabidopsis
15	299	14.0	695	5 Q9BHM3	Q9BHM3 paramecium
16	287.5	13.4	335	5 Q76719	Q76719 caenorhabdi
17	285	13.3	1166	10 Q9SYF6	Q9SYF6 arabidopsis
18	281	13.1	735	4 Q9UEQ5	Q9UEQ5 homo sapien
19	281	13.1	784	4 Q9UIE9	Q9UIE9 homo sapien
20	280	13.1	385	5 Q17909	Q17909 caenorhabdi
21	280	13.1	460	2 Q9ZIU2	Q9ZIU2 borrelia bu
22	279	13.0	369	16 Q83483	Q83483 treponema p
23	279	13.0	876	11 Q9WU62	Q9WU62 mus musculus
24	278.5	13.0	499	10 Q9FYW3	Q9FYW3 lycopersico
25	278	13.0	1108	5 Q9NDIO	Q9NDIO babesia big
26	277	12.9	513	10 Q9LW95	Q9LW95 nicotiana t
27	276	12.9	1339	11 Q35788	Q35788 rattus norv
28	275	12.9	451	16 Q50870	Q50870 borrelia bu
29	274.5	12.8	390	5 Q18401	Q18401 caenorhabdi
30	274	12.8	1062	5 Q960C4	Q960C4 drosophila
31	272	12.7	647	5 P91280	P91280 caenorhabdi
32	272	12.7	919	4 Q9NQS7	Q9NQS7 homo sapien
33	269.5	12.6	494	11 Q9JKL7	Q9JKL7 rattus norv
34	268.5	12.5	1750	3 Q8XOH2	Q8XOH2 neuropept
35	266	12.4	508	4 Q8WXA9	Q8WXA9 homo sapien
36	266	12.4	1052	5 Q26938	Q26938 trypanosoma
37	264.5	12.4	268	6 Q9SL36	Q9SL36 oryctolagus
38	264.5	12.4	312	5 P91570	P91570 caenorhabdi
39	264.5	12.4	810	10 Q9ZWI3	Q9ZWI3 cucurbita m
40	264	12.3	1430	5 Q9W0B0	Q9W0B0 drosophila
41	259	12.1	1432	10 Q23230	Q23230 arabidopsis
42	257.5	12.0	901	5 Q8SUW1	Q8SUW1 encephalito
43	257	12.0	546	4 Q9NW40	Q9NW40 homo sapien
44	256.5	12.0	1008	10 Q9FNE4	Q9FNE4 arabidopsis
45	256	12.0	873	13 Q13024	Q13024 xenopus lae

ALIGNMENTS

RESULT 1
Q8WWS6 PRELIMINARY; PRT; 763 AA.
ID Q8WWS6
AC Q8WWS6;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE DJ537K23.3 (KIAA0266 protein) (Fragment).
GN DJ537K23.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Grafham D.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL034405; CAD20048.1; -
FT NON_TER 763
SQ SEQUENCE 763 AA; 87075 MW; 41E3979FF06825C9 CRC64;

Alignment Scores:

Pred. No.: 5.74e-85 Length: 763
Score: 1338.50 Matches: 323
Percent Similarity: 63.62% Conservative: 32

Best Local Similarity: 57.89% Mismatches: 72
Query Match: 62.55% Indels: 133
DB: 4 Gaps: 8
US-09-502-945-3 (1-1298) x Q8WWS6 (1-763)

QY 11 ATGACTGCGAAGCGGCTTCAGAGAGCTTCGGCTTCAGCCANCAGGAAGAACTAGCG 70
DB 1 MetThAlaAsnArgLeuAlaGluSerLeuAlaLeuSerGlnGlnGluLeuAla 20
QY 71 GATTTCCCAAAAGACTTACCTCTGAGTGAGAGTGAAGATGAGGGGACAAATGATGAGAG 130
DB 21 AspLeuProLysAspTyrLeuLeuSerGluSerGluAspGluGlyAspAsnAspGlyGlu 40
QY 131 AGAAGCATCNAAGCTTCTGGAAGCAATCAGTTCCTTGTGATGGAAGAAATAGGGGAAA 190
DB 41 ArgLysHisGlnLysLeuLeuGluAlaLeuSerSerLeuAspGlyLysAsnArgArgLys 60
QY 191 TTGGCTGANAGGCTCAGCTAGTCTGAAGGCTCAGAGTTCATGCTCAATGCTGAGCA 250
DB 61 LeuAlaGluArgSerGluAlaSerLeuLysValSerGluPheAsnValSerSerGluGly 80
QY 251 TCAGGAGAAAAGCTGCTTCAGATCTGCTGAGCCTGTAAACTTCATCTTCTTGG 310
DB 81 SerGlyGluLysLeuValLeuAlaAspLeuLeuGluProValLysThrSerSerLeu 100
QY 311 GCCACTGTGAAAAGCACTGAGTAGAGTCANATCAAGAAANACAGTGGAGTTACCTCG 370
DB 101 AlaThrValLysLysGlnLeuSerArgValLysSerLysLysThrValGluLeuProLeu 120
QY 371 ACAAGAAGAGATTGAACGGATCCACAGAGATA-GCATTCAATAAAGCA-CAAGTC 428
DB 121 AsnLysGluGluLeuGluArgIleHisArgGluValAlaPheAsnLysThrAlaGlnVal 140
QY 429 CTCTCAATGGACCTGTCGCTCGCTGAAGACCGCAGCAGCAGCAGCTGTTTCC 488
DB 141 LeuSerLysTrpAspProValValLeuLysAsnArgGlnAlaGluLeuValPhePro 160
QY 489 CTGGAAGAGAGAGCCAGCTTCCATTTGACATGTGCTCAGTGGCTGGAAGGCA 548
DB 161 LeuGluLysGluGluProAlaIleAlaProIleGluHisValLeuSerGlyTrpLysAla 180
QY 549 AGAAGTCCCTGAGCAGGAGAAATTTCACTCTCCATGAAGCAAGCAGCCAGTGACA 608
DB 181 ArgThrProLeuGluGlnGluIlePheAsnLeuLysLysLysGlnProValThr 200
QY 609 GACCTTTACTGACCTGTGAAAAGGCTCTCTCCGAGCCATGAGCCTAGAGAGGCA 668
DB 201 AspProLeuLeuThrProValGluLysAlaSerLeuArgAlaMetSerLeuGluAla 220
QY 669 AAGATCGCAGCAGCAGCTTCAGAGGGCTCGGGCTCTGCAGTCCCTACTATGANGCCAAG 728
DB 221 LysMetArgArgAlaGluLeuGlnArgAlaArgAlaLeuGlnSerTyrTyrGluAlaLys 240
QY 729 GCTCGAAGAGAGAGAAATCNAAGTTAAAGTATACAAATCGTGAAGAAGAAAG 788
DB 241 AlaArgGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 260
QY 789 GCCAAGAAACCCCTAAAGAGTTGACAGCTCGGAGAGTTAATCCAGTCGCCGACCTA 848
DB 261 AlaLysLysAlaLeuLysGluPheGluGlnLeuArgLysValAsnProAlaAlaLeu 280
QY 849 GAAGACGAAGAAAGAGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 890
DB 281 GluGluLeuGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 300
QY 891 -----AAGAACAAG----- 899
DB 301 AsnSerGlyLysTrpAlaLysSerLysAlaIleMetAlaLysTyrAspLeuGluAlaArg 320
QY 900 -----GAGAAGAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 938
DB 321 GlnAlaMetGlnGluLeuSerLysAsnLysGluLeuThrGlnLysLeuGlnValAla 340

QY 939 GGAGAAGAGGAAAGGAAGAGGA----- 962
DB 341 SerGluSerGluGluGluGluGlyThrGluAspValGluGluLeuValProAsp 360
QY 962 ----- 962
DB 361 ValValAsnGluValGlnMetAsnAlaAspGlyProAsnProTrpMetLeuArgSerCys 380
QY 963 -----GAAAGAAAA 971
DB 381 ThrSerAspThrLysGluAlaAlaThrGlnGluAspProGluGlnLeuProGluLeuGlu 400
QY 972 GGAGAAGGAAAGGAAAA---CAAGGAGAAAGAAAGAACTAAGAAAGAGAGGAGAG 1028
DB 401 AlaHisGlyValSerGluSerGluGluGluArgProValAlaGluGluGluLeuLe 420
QY 1029 AATAAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1074
DB 420 uLeuArgGluPheGluGluArgArgSerLeuArgLysArgSerGluLeuSerGlnAspAla 440
QY 1075 -----GGAAGAGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1100
DB 440 aGluProAlaGlySerGlnGluThrLysAspSerGlySerGlnGluValLeuSerGluLe 460
QY 1101 NAGA-----AGAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1148
DB 460 uArgValLeuSerGlnLysLeuLysGluAsnHisGlnSerArgLysGlnLysAlaSerSe 480
QY 1149 AGAAGGA----- 1155
DB 480 rGluGlyThrIleProGlnValGlnArgGluGluProAlaProGluGluGluProLe 500
QY 1156 -----GAAGAAAGAAATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1202
DB 500 uLeuLeuGlnArgProGluArgVal-GlnThrLeuGluGluLeuGluGluGlyLysG 520
QY 1203 AAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1255
DB 520 luGluCysPheGlnAsnLysGluLeuProArgProValLeuGluGlyGlnGlnSerGluA 540
QY 1256 -----AGAAAGAAAGTATAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1291
DB 540 rgThrProAsnAsnArgProAspAlaProLysGluLysLysLysLysGlu 556

RESULT 2
Q9BVJ6 PRELIMINARY; PRT; 771 AA.
AC Q9BVJ6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Similar to KIAA0266 gene product.
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=CERVIX;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC001149; AAH01149.1;

RA Tanaka A., Kotani H., Miyajima N., Nomura N.:
 RT "Prediction of the coding sequences of unidentified human genes. VI.
 RT the coding sequences of 80 new genes (K1AA0201-K1AA0280) deduced by
 RT analysis of cDNA clones from cell line KG-1 and brain.";
 RL DNA Res. 3:321-329(1996).
 DR EMBL; D87455; BAA13396.1;
 SQ SEQUENCE 766 AA; 87187 MW; 4C90F0754A042986 CRC64;

Alignment Scores:
 Pred. No.: 3,41e-74 Length: 766
 Score: 1183.00 Matches: 289
 Percent Similarity: 61.22% Conservative: 52
 Best Local Similarity: 51.89% Mismatches: 84
 Query Match: 55.28% Indels: 134
 DB: 4 Gaps: 11

US-09-502-945-3 (1-1298) x Q92555 (1-766)

QY 11 ATGATCGGACCGCTTGCAGAGAGCTTGGCTTGGCCAGCCAGAGAACTAGCG 70
 Db 1 MetAsnValAsnGlnValAlaGluAsn---LeuAlaLeuSerHisGlnGluLeuVal 19
 QY 71 GATTGGCAAAAGACTACTCTTTCAGTGCAGAGTGAAGATGAGGGGACAAATGATGGAG 130
 Db 20 AspLeuProLysAsnTyrProLeuSerGluAsnGluAspGluGlyAspSerAspGlyGlu 39
 QY 131 AGAAGCATCNAAGCTTCTCGAAGCAATCATGTTCCCTTGTATGAAAGAAATAGCGGAAA 190
 Db 40 ArgLysHisGlnLysLeuLeuGluAlaIleLeuSerLeuAspGlyLysAsnArgArgLys 59
 QY 191 TTGCTGANAGTCTGAGGTCTGAGGTCTGAGGTCTGAGGTCTGAGGTCTGAGGTCTGAG 250
 Db 60 LeuAlaGluArgSerGluAlaSerLeuLysValSerGluPheSerValSerSerGluGly 79
 QY 251 TCAGAGAAAAGCTGCTTGCAGATCTGCTTGCAGCTGCTGTAACCTTCATCTCTTTG 310
 Db 80 SerGlyGluLysLeuGlyLeuAlaAspLeuLeuGluProValLysThrSerSerLeu 99
 QY 311 GCCACTGTGAAAGCAACTGAGTAGAGTCANATCAAGAAANACAGTGGAGTTACCTCTG 370
 Db 100 AlaThrValLysLysGlnLeuAsnArgValLysSerLysLysValValGluLeuProLeu 119
 QY 371 AACAAAGAGAGATGAGGATCCACAGAGAATA-CCATCAATAAAGCCCA-CAAGTC 428
 Db 120 AsnLysGluLysLeuGluGlnIleHisArgGluValAlaPheSerLysThrSerGlnVal 139
 QY 429 CTCCTCAAATGGACCTGCTGCTGCTGAAGAACCGGAGGAGCAGCAGCTGGTTTCCC 488
 Db 140 LeuSerLysTyrAspProIleLeuLysAsnGlnGlnAlaGluLeuValPhePro 159
 QY 489 CTGGAGAAAGAGGAGCCAGCCATGCTCCCATTTGAACATGTGCTCAGTGGCTGGAAGCA 548
 Db 160 LeuGlyLysGluGlnProAlaIleAlaProIleGluHisAlaLeuSerGlyTyrLysAla 179
 QY 549 AGAATCCCTCGGAGGAGAAATTTCAACCTCTCCATAGACACAGACAGCCAGTACCA 608
 Db 180 ArgThrProLeuGluGlnGluIlePheAsnLeuLeuHisLysAsnLysGlnProValThr 199
 QY 609 GACCTTTACTGACCCCTGTGGAAGGCGCTCTCTCCGAGCCATGAGCTAGAGAGGCA 668
 Db 200 AspProLeuLeuThrProMetGluLysAlaSerLeuGlnAlaMetSerLeuGluGluAla 219
 QY 669 AAGATGCCAGCAGAGAGCTTCAGAGGCTCGGGCTCGAGCTCGCTACTATGANGCCAA 728
 Db 220 LysMetHisArgAlaGluLeuGlnArgAlaArgAlaLeuGlnSerTyrTyrGluAlaLys 239
 QY 729 GCTCGAAGAGAGAAATCNAAGTTAAAGTATCAACAGTCGCAAGGCGCAAGAAAGAA 788
 Db 240 AlaArgLysGluLysLysIleLeuSerLysLysTyrHisLysValValLysLysGlyLys 259
 QY 789 GCCAAGAAAGCCCAAAAGAGTTTTCAGCAGCTGCGGAGGTTAATCCAGCTGCCGCACTA 848
 Db 260 AlaLysLysAlaLeuLysGluPheGluGlnLeuGlnLysValAsnProThrValAlaLeu 279

QY 849 GAAGAAGCAAGAAAG 890
 Db 280 GluGluMetGluLysIleGluAsnAlaArgMetMetGluArgMetSerLeuLysHisGln 299
 QY 891 -----AAGAACAAAG----- 899
 Db 300 AsnSerGlyLysTrpAlaLysSerLysAlaIleMetAlaLysTyrAspLeuGluAlaArg 319
 QY 900 -----GAGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 938
 Db 320 GlnAlaMetGlnGluGlnLeuAlaLysAsnLysGluLeuThrGlnLysLeuGlnValAla 339
 QY 939 GGAGAAG 971
 Db 340 SerGluSerGluGluGluGlyThrGluValGluLeuLeuValProHisVal 359
 QY 971 ----- 971
 Db 360 AlaAsnGluValGlnMetAsnValAspGlyProAsnProTrpMetPheArgSerCysThr 379
 QY 971 ----- 971
 Db 380 SerAspThrLysGluAlaAlaThrGlnGluAspProGluGlnValProGluLeuAlaAla 399
 QY 972 GGAGAAGAAAG 1031
 Db 400 HisGluValSerAlaSerGluAlaGluGluArgProValAlaGluGluLeuLeuLe 419
 QY 1032 AAGAAGGAAAG 1091
 Db 419 uArgGluPheGluGluGlnSerLeuArgLysArgSerGluLeuAsnGlnAspAlaG 439
 QY 1092 G-----GAAGAACTNAG 1103
 Db 439 uProAlaSerSerGlnThrLysAspSerSerSerGlnGluValLeuSerGluLeuAr 459
 QY 1104 A-----AGAAGAAAG 1151
 Db 459 gAlaLeuSerGlnLysLeuLysGlnLysHisGlnSerArgLysGlnLysAlaSerSerG 479
 QY 1152 AGGA-----GAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1164
 Db 479 uGlyThrValProGlnValGlnArgGluGluProAlaProGluGluAlaGluProLeu 499
 QY 1165 -ATAAG 1211
 Db 499 uLeuGlnArgSerGluArgValGlnThrLeuGluGluLeuGluGluLysGluAs 519
 QY 1212 G-----AAAAAG 1263
 Db 519 pCysPheGlnAsnLysGluLeuProArgProValLeuGluGlyGlnGlnSerGluArgTh 539
 QY 1264 -----AAGTATAAG 1287
 Db 539 rProAsnAsnArgProAspAlaProLysGluLysLysGluLysGluGln 555

RESULT 4
 O60531 PRELIMINARY; PRT; 131 AA.
 ID O60531
 AC O60531
 DT 01-AUG-1998 (TEMBLrel. 07, Created)
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE Antigen NY-CO-16.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-COLORECTAL CARCINOMA;
 RX MEDLINE-98272252; PubMed-9610721;

RESULT 10

Q9VYU0 PRELIMINARY; PRT; 2274 AA.

AC Q9VYU0; 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)

DE CG1905 protein.

DE CG1905.

GN Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Branton R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Balwle R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleeb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).

DR EMBL; AE003457; AAF48098.1;

DR FlyBase; FBgn0030337; CG1905.

SQ SEQUENCE 2274 AA; 247009 MW; B287311CBDF9BF20 CRC64;

Alignment Scores:

Pred. No.: 7,17e-14 Length: 2274

Score: 312.00 Matches: 128

Percent Similarity: 47.31% Conservative: 83

Best Local Similarity: 28.70% Mismatches: 136

Query Match: 14.58% Indels: 99

DB: 5 Gaps: 16

US-09-502-945-3 (1-1298) x Q9VYU0 (1-2274)

QY 101 AGTGAAGATGAGGGGACAAATGATGGAGAGAGAAAGCATCNAAGCTTCTCGAAGCAATC 160

|||||

Db 1203 SerGluAspGluAlaThr-----ThrThrIleThrLeuAlaLys 1215

QY 161 AGTTCCCTTGATGGAAAGATAGCGGAAATTTGGCTCGANAGGTTCTGAGGCTAGTCTCTCAAG 220

|||||

Db 1216 SerAlaLysLysAlaLysAsnAsnLysGln-----LysThrSerGlyLys 1230

QY 221 GTGTCAGAGTTCAATGTCAGTTCTGAAGGATCAGGAGAAAAGCTGGTCTCTTCGACATCTG 280

|||||

Db 1231 ValSerGluLysLysProGluGlnThrAsnLeuAlaGluGluLysVal----- 1246

QY 281 CTTGAGCCTGTAAACATCTCATCTTCTGGCCACTCTGAAAAGCAACTGAGTAGATC 340

|||||

Db 1247 -----AlaValGluLysGluGluAsnGluGlu 1255

QY 341 ANATCAAGAANACAGTGGAGTTACCTCTGAACAAAGAA-----GAGATTGAACGG 391

|||||

Db 1256 ArgAlaValProMetValLysLysThrIleAsnLysGluGluAspSerAspAlaGluSer 1275

QY 392 ATCCACAGAGATAGCATTCATTAATAACGACGACAGTCTCTCCAAATGGGACCTGTCTG 451

|||||

Db 1276 AsphHisAlaAsp-SerLeuLeuAlaAsnLysSerSerIleAla-----AlaValMe 1292

QY 452 CCTGAAGAACCGGCGACGACGAGCTGGTTTTCCTCCGTGAGAAAAGAGAGCCGCCAT 511

|||||

Db 1292 tValSerSerAlaSerAlaGlnGlyLeuSerLeuHisValGlu----- 1306

QY 512 TGTCCCATTTGAACATGTGCTGCTGCTGGAAGGCAAGAACTCCCTCGAGCAGCAAAAT 571

|||||

Db 1307 -----MetSerAlaAlaAspAlaGluGlnGlyLeuAspGluGlu 1320

QY 572 TTTCAACTC-----CTCCATAAGAACCAAGCAGCC 601

|||||

Db 1320 eGluGlyLeuAspGluGluProLysThrMetSerLysAspAsnLysLysLysGln 1340

QY 602 AGTGACAGACCTTTACTGACCCCTGTG-----GAAAGGCCCTCTCTCCG 646

|||||

Db 1340 sProGlyAspAlaValAlaThrMetThrIleAspLysGluLysGluLysAlaLysGlu 1360

QY 647 AGCCATGAGCTAGAGAGGCAAGATGCGACGAGCAGAGCTTCAG----- 692

|||||

Db 1360 sGluLeuLysLysGlu-----LysGluArgGluAlaLysLeuGlnGluLysGluLysGlu 1379

QY 693 -----AGGCTCGGGCTCT 706

Db 1379 uGluLysLysLysLysLysGluArgGluGluSerLeuArgMetGluArgGluGluLysLe 1399

QY 707 GCAGTCTCTACTATGANGCCCAAGCTCGAAGAGAGAGAAATCNAAGTTTAAAGTATCA 766

|||||

Db 1399 uLysGluGluLysIleLysGluLysGluArgGluGluGluLysLeu-----LysGluGlu 1416

QY 767 CAAAGTCGTGAAGAAAGGCAAGCCCAAGAAAGCCCTTAAAGAGTTTGACAGCTCGGAA 826

|||||

Db 1416 uLysIleLysGluLysGlnArgGluGluLys-----LeuLysGlu-----GluLysLysLysGlu 1434

QY 827 GGTATATCCAGCTGCCCATCTAGAACACCAAGAAAGAGGAAAGA-----AGGAG 877

|||||

Db 1434 uLysGluArgGluGluArgMetLysGluLysGluArgGluGluLysAlaLysGluLysGlu 1454

QY 878 GAGGAGAAAGCAAGAAAGCAAGGAGAGAAAGAAAGGAGGAGGAGGAGGAGGAGAAAGA 937

|||||

Db 1454 nArgGluGluLysLeuArgGluLysIleLysGluLysGluLysGluGluGluGluLysLeu 1474

QY 938 AGGAGAGAGAGAAAGGAG 985

|||||

Db 1474 yGluLysLeuArgGluGluLysIleLysGluLysGluLysGluLysGluLysLysGluLysGlu 1494

QY 986 AAAAGAGGAG 1045

|||||

Db 1494 luArgGluGluLysMetArgGlu--LysGluArgGluGluLysIleLysGluLysGluLysGlu 1513

QY 1046 AAAGAAAGAGTAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1105

|||||

Db 1513 gValGluLysIleLysGluLysGluArgGluGluLysLysLysGluLysGluLysGluLysGlu 1533

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-T., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPRAIN-BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF003530; AAF49613.2;
 DR EMBL: AY069576; AAL39721.1;
 DR Flybase; FBgn0036514; CG12301.
 DR SEQUENCE 771 AA; 88727 MW; 912B33FE58D7CEDD CRC64;
 SQ

Alignment Scores:
 Pred. No.: 1-13e-13 Length: 771
 Score: 309.50 Matches: 123
 Percent Similarity: 45.15% Conservative: 77
 Best Local Similarity: 27.77% Mismatches: 180
 Query Match: 14.46% Indels: 63
 DB: 5 Gaps: 15

US-09-502-945-3 (1-1298) x Q9VUR2 (1-771)

QY 101 AGTGAAGATGAGGGGACAAATGATGGAGAGAGAAAGCATCAAGCTCTCGAAGCAATC 160
 DB 1 SerAspAspGlu---GluHisTyrGluProLysAlaHisLysLysLeuLeuGlnAlaIle 20
 QY 161 AGTTCCTTCATGGAAGAATAGCGGAATGGCTGACAGAGTCTGAGGCTAGTCTGAAG 220
 DB 21 GlySerGluGlyLysValGlnHisIleGlnLysSerThrArgAspGluArgGlnSerLeu 40
 QY 221 GTCTCAGAGTCTC-----AATGTCAGTCTGAGGATCAGGAGAA 259
 DB 41 GlnAspGluPheGlnLeuValLysGlyValSerSerAlaGluThrAspHisAlaProArg 60
 QY 260 AACTGCTCTCGAGATCTGTGAGCCCTGTAAACTTCTCTTCTTTGGCCACTGTG 319
 DB 61 AlaValGlyLeuAsnAspLeuValAspIleLeuArgThrSerThrLysHisSerGlnThr 80
 QY 320 AAAAAGCAACTGAGTAGATC---ANATCAAGAAACAGTGGGTACCTCTGACACAA 376
 DB 81 GlyLysLysLysLysAsnIleHisGlySerLysValLysValLysProLeuGluLys 100
 QY 377 GAAGAGATTGAACGGATCCACAGAGATAGCATCAATAAAGCAAGCAAGTCTC----- 428
 DB 101 ProAlaAlaAspArg-----IleLysArgThr-IleGlyTyrGln 113
 QY 429 -----CTCTCAATGGGACCTGTGCTGCTCTGAGAACCGGACGAGCAGA 472
 DB 113 uGlyValThrLysLysLysLeuGlyArgTrpAspAlaValAlaGlnGlnArgSerAlaGln 133
 QY 473 GACGCTGGTTTCCCTCGAGAAAGAGGAGCCAGCATGCTCCCATGGAACATGTGCT 532
 DB 133 uThrGlnIlePheProLeuProSerGluThrValTyrValAsnThrAlaAlaAsnAlaAr 153
 QY 533 CAGT---GGCTGGAAGGCAAGAACTCCCTGGAGCAGGAAATTT-----TTCAA 577
 DB 153 gProLeuAsnThrArgValLysSerAsnLeuAlaGlnGluLeuAlaAsnAsnArgLy 173
 QY 578 CTTCTCTCCATFAGAACAGCAGCAGTGTGACAGCCCTTTA-----CTGACCCC 625
 DB 173 sLeuArgGluLeuArgLysAlaGlnIleGlyAspThrThrAspGluLysGluLeuAlaLy 193

QY 626 TGTGGAAAAGGCTCTCTCTC---CGAGCCATGACCTAGAACGAGGCAAGATCGCAGCAGC 682
 DB 193 sGlnGluArgLeuLeuGlnLysLysLeuThrArgAspGluLeuPheAlaArgArgLy 213
 QY 683 AGAGCTTCAGAGGCTCGGGCTCTGCAGTCTCTACTATGANGCCCAAGCTCGAAGAGAGAA 742
 DB 213 sGluLeuAlaTyrLeuLysMetArgGlnSerGlnLysSerAlaLysAlaArgMetGlnAs 233
 QY 743 GAAATCNAAAGTTAAAGTATCACAAGTCGTGAAGAAGGAAAGCCCAAGAAAGCCCT 802
 DB 233 nLysIleLysSerLysLysPheHisLysLysLeuGlnLysArgGlnLysMetLeuGlnMe 253
 QY 803 AAAAGAGTTTGCAGCTCGCGAGGTTTATCCAGCTGCC-----GC 844
 DB 253 tLysGluPheGluLeuGlnLysThrAspProGluAlaAlaLeuGluLysLeuAsnAl 273
 QY 845 ACTAGAGAAGCAAGAAAG 904
 DB 273 aLeuGluLysSerArgValGlnGluArgAlaSerLeuArgHisLysAsnThrGlyThrTr 293
 QY 905 GAAGAAGAAG 964
 DB 293 pAlaLysAsnLeuGlnIleArgAlaLysTyrAspLysAspValArgLysAspLeuAlaGln 313
 QY 965 AGAAAAG 1024
 DB 313 uGlnLeuAlaValSerArg-----GluLeuThrGlnLysLysGlnValSe 328
 QY 1025 GAAGAATAAG 1078
 DB 328 rAspSerGluAspGluThrLysArgValLeuProGluGluGluAspAsnAspTy 348
 QY 1079 GAAGAAAG 1138
 DB 348 rAspProPheAsnProTrpThrLysArgLysAlaThrGluGlnThrGluAsnGlyAsnGln 368
 QY 1139 GGAACNAGAA-----AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1174
 DB 368 yGluSerGluGlySerAsnTrpArgGlnTyrTrpThrLysArgAsnLysAsnGluLysLe 388
 QY 1175 GAAGAAGAAAG 1216
 DB 388 uLeuGlu-GluHisArgArgAspLeuGluGlyValGlnAspAsnAspGluGlnGluLul 408
 QY 1217 AGCAAGAAAG 1276
 DB 408 ySgluThrIleArgLysProLeuIleGluValLysProAlaLysLysLysLysLys 428
 QY 1277 AAGAA 1281
 DB 428 erLys 429
 RESULT 13
 Q8T5C7 PRELIMINARY; PRT; 2055 AA.
 ID Q8T5C7
 AC Q8T5C7
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE MAEBL.
 GN MAEBL.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Blair B.L., Kappe S.H.I., Maciel J.E., Balu D.J., Adams B.;
 RT "Plasmodium falciparum maeli is a unique member of the ebl family."
 RL Mol. Biochem. Parasitol. 0:0-0(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Michon P., Stevens J.R., Kaneko O., Adams J.H.;
 RT "Plasmodium falciparum maeli is a unique member of the EBL family.";

Search completed: March 21, 2003, 12:58:56
Job time : 64.5135 secs

```
QY 354 CAGTGGAGTTACTC-----TGAACAAGAAGAGAGATTGAACGGATCCACAGAG 401
||||| |||
Db 307 GlnTrpSerLeuLeuLysLysAlaGlnArgLysLysLysGluLeuAsnLeuGlnArgGlu 326
||||| |||
QY 401 ----- 401

Db 327 LeuAlaGlnArgSerLysAsnLysAspHisSerPheIleGluSerLeuLeuLysLys 346

QY 402 -----AATAGCATTCAATAAAAGCCACACAGTCTCTCCAAA 437
||||| |||
Db 347 AspTyrPheHisThrLysAlaAsnAspPheValIleAspGlnGlnLeuGlnLys 366
||||| |||
QY 438 TGGGACCTGTCTCTGAAGAACCGGACGAGCAGCTGTTTTCCTCCCTGGAGAAA 497
||||| |||
Db 367 -----LeuGlnAspGlnGlnLeuLeuGlnMetGlnLeuAspSer 381
||||| |||
QY 498 GAGGAGCCACCATCTCCCATGTAACATGTCTCAGTGGTGGAGGCAAGACTCCC 557
||||| |||
Db 382 GlnGlnLysIle-----GlnGlnGlnThrGln 390
||||| |||
QY 558 CTGAGCAGAGAAATTTTCAACCTCTCCATAGAACAAGCAGCCAGTGCAGACCCCTTTA 617
||||| |||
Db 391 LeuLysGlnGluLysGlnLysLysGluLysLysAsnLysLys----- 404
||||| |||
QY 618 CTGACCCCTGTGAAAGGCTCTCTCCGAGCCATGAGCCTAGACAGGCAAGATGCCA 677
||||| |||
Db 405 ---LysSerAlaGluLysLysAsnLysArgHisArgSerSerGlnGlyGluSerAsp 423
||||| |||
QY 678 CGAGCAGAGCTTCAGAGGCTCGGGCTCTGAGTCTCTACTATGANGCCAGGCTCGAAGA 737
||||| |||
Db 424 -----SerAspSerGlnSerLeuLysasn-----LysLysGlnLys 435
||||| |||
QY 738 GAGAAGAAATCAAAAGTTAAAGTATACAAAGTCTGTAAGAAAGGAAGGCCAAGAAA 797
||||| |||
Db 436 AsplHisLysLysAsnLysHisLysSerTyrAspArgAspLysAspGlnGln 455
||||| |||
QY 798 GCCTTAAGACAGTTTGAGCAGCTGCGGAGGTTAATCCAGCTCCGCTAGAGAACGA 857
||||| |||
Db 456 LysArgLysGluLysGluLysLysLysGlu-----GlnGluLys 468
||||| |||
QY 858 AGAAAGAGGAGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 917
||||| |||
Db 469 GlnGlnAspLeuGluLysGlnLysGluLysGlnLysGlnLysGlnLysGlnGln 488
||||| |||
QY 918 GGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 977
||||| |||
Db 489 GluLysGluArg-GluLysGluLysGluLysGluLysGluLysGluLysGluLys 508
||||| |||
QY 978 GGAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1037
||||| |||
Db 508 gGluLysGlu---ArgGluLysGluLysGluLysGluLysGluLysGluLysGluLys 526
||||| |||
QY 1038 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1097
||||| |||
Db 527 GluLysGluArgGluArgLysArgAspArgGluArgGluLysGluLysGluArgLeuArg 546
||||| |||
QY 1098 ACTNAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1157
||||| |||
Db 547 GluLysGluLysGlnArgGluLysAspArgGluArgGluLysGluLysGluLysArg 566
||||| |||
QY 1158 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1217
||||| |||
Db 567 LysArgGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 586
||||| |||
QY 1218 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1277
||||| |||
Db 587 GlyLysGlnGluGlnLysLeuLysGluLysGluLysGluLysGluLysGluLys 603
||||| |||
QY 1278 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298
||||| |||
Db 604 GlyLysAsnLysGlnSerGlu 610
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 08:38:27 ; Search time 54.6146 Seconds
(without alignments)
10910.938 Million cell updates/sec

Title: US-09-502-945-4

Perfect score: 4120

Sequence: 1 cctggcccggtcggtcgccg.....tattttccagcttaaaaaa 2236

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O/cgn2_1/USPTO_SPOOL/US09502945/runat_14032003_101057_19087/app_query.fasta_1.10979
-DB-A_Geneseq_101002 -OEMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_365 -runat_14032003_101057_19087 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	3371	81.8	652	20	AY07094
2	3371	81.8	652	20	AY07050
3	2331	56.6	481	20	AY07098
4	2311	56.1	521	20	AY07099
5	2046	49.7	403	20	AY07095
6	1732	42.0	328	20	AY07097
7	1350	32.8	268	20	AY38440
8	447	10.8	120	21	AA53328
9	425.5	10.3	493	22	AB861924
10	417	10.1	98	20	AY38390
11	406	9.9	95	20	AY38437
12	272	6.6	1037	22	ABG22366
13	271.5	6.6	928	23	AA211718
14	267.5	6.5	2037	21	AA533753
15	266.5	6.5	1111	23	AA087918
16	263.5	6.4	767	23	AA084269
17	263.5	6.4	767	23	AA084328
18	261.5	6.3	767	21	AA22138
19	261.5	6.3	1745	23	AB57253
20	261	6.3	1535	23	AA087934
21	259.5	6.3	856	21	AA01383
22	259.5	6.3	1881	20	AA24025
23	258.5	6.3	724	23	AB04798
24	258.5	6.3	1526	22	AA07977
25	256.5	6.2	1526	22	ABG06116
26	256.5	6.2	1526	22	ABG07289
27	256.5	6.2	1552	22	AA07893
28	256.5	6.2	1554	22	ABG07288
29	256.5	6.2	1584	22	ABG07290
30	256.5	6.2	1627	22	ABG06117
31	254.5	6.2	2070	20	AY04733
32	247.5	6.0	763	20	AY04741
33	247.5	6.0	1005	20	AY04731
34	247.5	6.0	1373	20	AY04730
35	247.5	6.0	2000	20	AY04732
36	242.5	5.9	724	23	AB04805
37	237	5.8	46	22	AB38777
38	237	5.8	46	22	AB23814
39	237	5.8	46	22	AA59416
40	237	5.8	46	22	AA071974
41	237	5.8	46	22	AA019386
42	237	5.8	46	22	AA032249
43	237	5.8	46	23	ABG41789
44	234.5	5.7	817	19	AA048101
45	234.5	5.7	817	20	AA030137

ALIGNMENTS

RESULT 1
AA07094
ID AA07094 standard; Protein; 652 AA.
XX
AC AA07094;
XX
DT 02-JUL-1999 (first entry)
XX
DE Colon cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
OS Homo sapiens.
XX
PN W09904265-A2.
XX
PD 28-JAN-1999.


```

XX PF 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX WPI: 1999-132448/11.
XX PR New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX PS Disclosure; Page 656-658; 787pp; English.
XX CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX SQ Sequence 652 AA;

Alignment Scores:
Pred. No.: 2,75e-276 Length: 652
Score: 371.00 Matches: 652
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.82% Indels: 0
DB: 20 Gaps: 0

US-09-502-945-4 (1-2236) x AAY07094 (1-652)
QY 97 ATGACCGAAGACATCTCTATGATGTCGCGAATGACACAGACCATGGACGTGGATTTCTGATTGAAATGAT 156
Db 1 MetAspArgLysValAlaArgGluPheHisLysValAspPheLeuLeuGluAsnAsp 20
QY 157 GCAGAGAAGACATCTCTATGATGTCGCGAATGACACAGACCATGGACGTGGCC 216
Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
QY 217 GTGCTCTGGGAGACCTGAAGCTGGTCATCAATGAACCCAGCCGCTCTGCTCTGTTGAT 276
Db 41 ValLeuValGlyAspLeuLysLeuValAlaAsnGluProSerArgLeuProLeuPheAsp 60
QY 277 GCATTCGGCCGCTGATCCACCTGAAGACACCATGGATATGATCAGCTGACCCCGG 336
Db 61 AlaIleArgProLeuLeuProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
QY 337 CGCTCAGGAAGCTGAAGAGGTGCGTCTGGACCGCTGACACCCGAGCCCTGCGCTG 396
Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
QY 397 AGTGTGGCTGGCTGGAGTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGGC 456

```

```

Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheHisSerHisLeuLeuLysGly 120
QY 457 GGTCAAGCAGACAGCGTCGGCTCCAGGTAGGGACGAGATCGTCCGATCAATGGATAT 516
Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluLeuValArgIleAsnGlyTyr 140
QY 517 TCCATCTCTCTGTACCCATGAGAGGTGATCAACCTCATCTGACCAACCAAACTGTG 576
Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
QY 577 TCCATCAAGTGAGACATCGGCTGATCCCGTGAAGCTCTCTCATGATGAGCCCTC 636
Db 161 SerIleLysValArgHisIleGlyLeuLeuProValLysSerSerProAspGluProLeu 180
QY 637 ACTTGGCAGTATGTGGATCAGTTTGTGCGAATCTGGGGCGTGCAGGAGCCTGGGC 696
Db 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlySerLeuGly 200
QY 697 TCCCTCGAATCGGAAACAGAGAGAGAGGTCTTCATCAGCCTGGTAGGCTCCGA 756
Db 201 SerProGlyAsnArgGluAsnLysLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCCTTGGCTGCAGCATTTCCAGCGGCCCATCCAGAGCCTGGCATCTTTATCAGCAT 816
Db 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTGAACCTGGTCCCTCTGCTGAGTGGGATGGAGTAGGAGGACCAAGATTCGAA 876
Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCAATGGCTGCTACTTCTTAACCTGCATCAAGAGGCTGTAAATGCTGCTGAAAT 936
Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
QY 937 ACCCCAGCCTCAGCATCTCCATTTAGCTGCAGCTGGCCGGAGCTGTTCATGACAG 996
Db 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGAGCGGCTGGCAGAGCGCGCTGAGCTGAGCTGCAGCGCAGAGCTTCTCATGCG 1056
Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuLeuMetGln 320
QY 1057 AAGCGCTGGCGATGAGTCCCAACAGATCTCCAGGAGCAGCAGAGATGGAGCGCAA 1116
Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGluGlnGlnGluMetGluArg 340
QY 1117 AGGAGAAAGAAATTCGCCAGAGGCAGCAGAGGAAATAGAGATACCGAAGGAGATG 1176
Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
QY 1177 GAACAGATTGTAGAGAGAGAGAGAGAGTTTAAAGACCAATGGGAAGAGACTGGGGCTCA 1236
Db 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluLysArgLysGlySer 380
QY 1237 AAGGAACAGTACTCTGCTTAAACCATCACTGCTGAGGTACACCCAGTACCCCTCGC 1296
Db 381 LysGluGlnLeuLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCCAAAGTATGATCAGGAGTGGAACTGAGCTCGAGCCCGCAGATGACCTGGATGA 1356
Db 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspAspLeuAspGly 420
QY 1357 GGCAGGAGGAGCAGGAGAGAGATTTCCGGAATATGAGGAAGGCTTTCACCCCTAC 1416
Db 421 GlyThrGluGluGlnGlyGluGlnAspPheArgLysTyrGluGluGlyPheAspProTyr 440
QY 1417 TCTATGTTCAACCCAGACAGATCATGGGAGGATGTCCGCTCTCTACGCATCAAGAG 1476
Db 441 SerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLys 460
QY 1477 GAGGATCTTACAGCTGGCCCTGGAAGCGGTGGAGTCCCGCATTTGGGAAGGTGTC 1536
Db 461 GluGlySerLeuAspPheAlaLeuGluGlyValAspSerProIleGlyLysValVal 480

```


QY 1537 GTTCTGCTGTATGAGCGGGAGCTGCTGAGCGGATGCGTGCATTGTGAAGGGGAC 1596
Db 481 ValSerAlaValTYrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAsp 500
QY 1597 GAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCCCTGGCTGAGGCTGACGCT 1656
Db 501 GluIleMetalIleAsnGlyLysIleValThrAspTYrThrLeuAlaGluAlaAspAla 520
QY 1657 GCCCTGCAGAGGCTGGAATCAGGGCGGGGACTGAGTCGACCTTGTTGGTGGCTGCTGC 1716
Db 521 AlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValAlaValLys 540
QY 1717 CCCCCAAGAGTATGACGATGACCTGACCTTCTGCTGAGTCCAAAGGGGAAACAA 1776
Db 541 ProProLysGluTYrAspAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnGln 560
QY 1777 ATTCACCGCTTAGCAACAGTAGCTCCGCCCCACCTCGTGACACAAAGCCCTCGGACC 1836
Db 561 IleHisAlaLeuGlnSerGluLeuArgProHisLeuValAsnThrLysProArgThr 580
QY 1837 AGCCTTGAGAGGCCACATGACACACACAGATGGCATCTCTGGACCTGAATCTATCA 1896
Db 581 SerLeuGluArgGlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSer 600
QY 1897 CCGAGGAATCTCAACTCCCTTTGGCCCTGAAACCAGGGCCAGATPAGGAACAGCTCGGGC 1956
Db 601 ProArgAsnLeuLysLeuProLeuAlaLeuAsnGlnGlyGlnIleArgAsnSerSerGly 620
QY 1957 CACTTTTGAAGCCCAATGAGGAGGAAGGGAGGACCCAGCCGTTTGGGAGAGATCTC 2016
Db 621 HisPheGluGlnGlnCysGlyGlyLysGlyAlaAlaSerArgLeuGlyGluAspLeu 640
QY 2017 AAGGATCCAGACTCTCATCTCTTCTCTCTGCCCCAG 2052
Db 641 LysAspProAspSerHisSerPheProLeuAlaGln 652
RESULT 2
AAV07050.
ID AAV07050 standard; Protein; 652 AA.
XX AC AAV07050;
XX DT 02-JUL-1999 (first entry)
XX DE Renal cancer associated antigen precursor sequence.
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX OS Homo sapiens.
XX PN WO9904265-A2.
XX PD 28-JAN-1999.
XX PF 15-JUL-1998; 98WO-US14679.
XX PR 23-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
XX PR 10-OCT-1997; 97US-0061599.
XX PR 10-OCT-1997; 97US-0061765.
XX PR 10-OCT-1997; 97US-0948705.
XX PR 11-OCT-1997; 97GB-0021697.
XX (LUDW-) LUDWIG INST CANCER RES.
XX PA Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX WPI; 1999-132448/11.
DR

XX

PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers

XX PS Disclosure; Page 436-438; 787pp; English.

XX CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

SQ Sequence 652 AA;

Alignment Scores:

Pred. No.: 2,75e-276 Length: 652
Score: 3371.00 Matches: 652
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.82% Indels: 0
DB: 20 Gaps: 0

US-09-502-945-4 (1-2236) x AAV07050 (1-652)

QY 97 ATGGACCGAAAGTGGCCGAGAAATTCGGCATAGTGGATTCTGTGTTGAAATGAT 156

Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20

QY 157 GCAGAGAAGGACTATCTCTATGATGCTGCGAATGTACACACAGACCATGACGTGGCC 216

Db 21 AlaGluLysAspTYrLeuTYrAspValLeuArgMetTYrHisGlnThrMetAspValAla 40

QY 217 GTGCTCGTGGGACACTGAAGCTGTCATCAATGAACACCGCTGCTGCTCTGTTGAT 276

Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60

QY 277 GCCATTGCGCCGCTGATCCACTGAAGCACAGGTTGGAATATGATCAGCTGACCCCGG 336

Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTYrAspGlnLeuThrProArg 80

QY 337 CGCTCCAGGAAGCTGAAGAGGTGCTGTCGACCGCTGCACCCGAGGCTCGGCCCTG 396

Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100

QY 397 AGTGTGCGTGGTGGCTGGAGTTGGCTGTGGGCTTTCATCTCCACATCATCAAGGC 456

Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120

QY 457 GGTACGGACAGACGCTCGGGCTCCAGGTAGGGGACGAGATCGTCCGGATCAATGGATAT 516

Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTYr 140

QY 517 TCCATCTCTCTGTACCCATGAGGAGGTTCATCACTCATTCGACCAAGAAACTGTG 576

Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160

QY 577 TCCATCAAGTGAACACATCGGCCCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTC 636

Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180

QY 637 ACTTGGCAGTATGTCAGTTTGTCTGCGAATCTGGGGGCTGCGAGGACGCTCGGC 696

Db 181 ThrTrpGlnTYrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200

QY 697 TCCCTCGAATCGGAAACACAGAGAAAGGTCTTTCATCGCTGGTAGCTCCCGA 756
|||||
Db 201 SerProGlyAsnArgGluAsnLysGluLysLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCTTGGCTGACGATTTCCACGGCGCCCATCCACAGCCTGGCATCTTTATCAGCCAT 816
|||||
Db 221 GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTCAAACTGGCTGCCCTGCTGCTGAGTGGGATTGGAGATAGGGACACAGATTGTCGAA 876
|||||
Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCATAGGCTGCATCTCTTAACCTGGATCACAAGAGGCTGTAATGCTGGAATAAT 936
|||||
Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
QY 937 AGCGGAGCTGACCATCTCATTGTAGCTGCAGCTGGCGGAGCTGTTCATGACAGAC 996
|||||
Db 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGAGCGCTGCGAGCGCGCGCGCTGAGCTGCAGCGCAGGAGCTTCTCATCGAG 1056
|||||
Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuMetGln 320
QY 1057 AAGCGCTGCGATGGAGTCCACAGATCTCCAGGAGCAGCAGGATGGAGCGCAA 1116
|||||
Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln 340
QY 1117 AGAGAAAGAAATGCGCAGAGAGCAGCAGAGAGAAATGAGAGATACCGAAGGAGATG 1176
|||||
Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGlnLysGlnArgLysGluMet 360
QY 1177 GAACAGATTGTAGAGGAGGAGAGAGTAAAGAACCAATGGGAAGAGCTGGGCTCA 1236
|||||
Db 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluLysTrpGlySer 380
QY 1237 AAGAACAGCTACTTCTGCTTAAACATCAGCTGAGTACACCCAGTACCCCTCGC 1296
|||||
Db 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCAAAGTATGATCAGGAGGTGAACCTGAGCTCGAGCCCGCAGATGACCTGGATGA 1356
|||||
Db 401 LysProLysTrpAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
QY 1357 GGCACGGAGGAGCAGGAGCAGGATTTCCGGAATATATAGGAAGGCTTTGACCCCTAC 1416
|||||
Db 421 GlyThrGluGluGlnGlyGluAspPheArgLysTrpGluGluGlyPheAspProTyr 440
QY 1417 TCTATGTCACCCAGACAGATCATGGGAAGGATGTCGGCTCTCTACCATCAAGAAG 1476
|||||
Db 441 SerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLys 460
QY 1477 GAGGATCTTACACCTGGCCCTGGAGGCGGTGTGACTCCCCCATGGGAAGGTGTC 1536
|||||
Db 461 GluGlySerLeuAspLeuAlaLeuGluGlyGlyValAspSerProIleGlyLysValVal 480
QY 1537 GTTCTCTCTGTATGACGGGAGCTGCTGACGGGATGTCGCGCATGTGGAAGGGAC 1596
|||||
Db 481 ValSerAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAsp 500
QY 1597 GAGATCATGCAATCAACGGCAAGATTGTGACAGACTACACCTGGCTGAGGTGACGCT 1656
|||||
Db 501 GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAla 520
QY 1657 GCCCTCAGAGCGCTGGAATCAGGGCGGGAGCTGATGACCTGTTGGTGGCTGCTGC 1716
|||||
Db 521 AlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValValAlaValCys 540
QY 1717 CCCCCAAGGAGTATGACGATGAGCTGACCTCTCTGCTGAAGTCCAAAGGGGAACCAA 1776
|||||
Db 541 ProProLysGluTyrAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnGln 560

QY 1777 ATTACGCGTTAGGAACACAGTGAAGTCCGGCCCGCCACCTCGTGAACACAAAGCCTCGGACC 1836
|||||
Db 561 IleHisAlaLeuGlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgThr 580
QY 1837 AGCCTTGAGAGAGGCGCATGATCACACACACAGATGGCATCTTGGGACCTGAATCTATCA 1896
|||||
Db 581 SerLeuGluArgGlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSer 600
QY 1897 CCCAGGAATCTCAAACTCCCTTGGCCCTGAACACGAGCCAGATGAAGAACAGCTCGGCG 1956
|||||
Db 601 ProArgAsnLeuLysLeuProLeuAlaLeuAsnGlnGlyGlnIleArgAsnSerSerGly 620
QY 1957 CACTTTTGAAGGCCAATGTGGAGGAAGGAGGAGCAGCAGCCGTTTGGGAGAGATCTC 2016
|||||
Db 621 HisPhePheGluGlnCysGlyGlyLysGlyAlaAlaSerArgLeuGlyGluAspLeu 640
QY 2017 AAGGATCCAGACTCTCATCTCTTCTCTGCGCCAG 2052
|||||
Db 641 LysAspProAspSerHisSerPheProLeuAlaGln 652
RESULT 3
AA07098
ID AA07098 standard; Protein; 481 AA.
XX AC AA07098;
XX DT 02-JUL-1999 (first entry)
XX DE Colon cancer associated antigen precursor sequence.
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX OS Homo sapiens.
XX PN WO9904265-A2.
XX PD 28-JAN-1999.
XX PF 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX WPI; 1999-132448/11.
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX Disclosure: Page 664-665; 787pp; English.
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.

XX
 SQ Sequence 481 AA;

Alignment Scores:

Pred. No.: 2,41e-188 Length: 481
 Score: 2331.00 Matches: 462
 Percent Similarity: 96.06% Conservative: 1
 Best Local Similarity: 95.85% Mismatches: 0
 Query Match: 56.58% Indels: 19
 DB: 20 Gaps: 1

US-09-502-945-4 (1-2236) x AAY07098 (1-481)

QY 97 ATGGACGAAAGTGGCCGAGAAATCCGGCATAAAGTGGATTCTTGATGAAATGAT 156
 DB 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuLeuGluAsnAsp 20
 QY 157 GCAGAGAAGACTATCTCATGATGTGCTCGGAATGTACCACAGACCACCTGGACGTGGCC 216
 DB 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
 QY 217 GTGCTCGTGGAGACCTGAAGCTGGTCATCAATGAACACAGCGCTGCTCTGTTTGTAT 276
 DB 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
 QY 277 GCCATTCCGGCGCTGATCCACTGAAGCACACAGGTGGAATATGATCAGCTGACCCCGG 336
 DB 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 QY 337 CGTCCAGGAAGCTGAAGAGGTGCTGTGGACCGTCTGCACCCGAGGCCCTCGGCCCTG 396
 DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgHisProGluGlyLeuGlyLeu 100
 QY 397 ACTGTGCTGGTGGCTGGAGTTGGCTGTGGCTCTTATCTCCACCTCATCAAGGC 456
 DB 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuLysGly 120
 QY 457 GGTACGACAGACGCTCGGGCTCCAGGTAGGGAGCAGATCTCCGGATCAATGATAT 516
 DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 QY 517 TCCATCTCTCTGTACCATGAGGAGTTCATCACTCATTCGAACCAAGAACTGTG 576
 DB 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysThrVal 160
 QY 577 TCCATCAAAAGTGAGACACATCGCCGTGATCCCGTGAAAGCTCTCTGATGAGCCCTC 636
 DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
 QY 637 ACTTGGCAGTATGTGGATCAGTTGTGTCGAATCTGGGGGTGCGAGGACGCTGGGC 696
 DB 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200
 QY 697 TCCCTTGGAAATCGGAAACAGGAGAGAGTCTTCATCAGCCTGTAGCTCCCGA 756
 DB 201 SerProGlyAsnArgGluAsnLysGluLysLysValPheIleSerLeuValGlySerArg 220
 QY 757 GGCTTGGCTGCGACATTTCCAGCGGCCCCATCCAGACCTGGCATCTTTATCAGCCAT 816
 DB 221 GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
 QY 817 GTGAACCTCGGCTCCCTGCTGCTAGGTGGGATGGAGATAGGGACAGAGTTGTCGAA 876
 DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
 QY 877 GTCAATGGCGTGCACCTTCTTAACCTGGATCAAGAGGCTGTAAATGTCTGTAAGAAAT 936
 DB 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280

QY 937 AGCCGAGCGCTGACCATCTCCATTGTAGTGTGCGAGTGGCCGGAGCTGTTCATGACAGAC 996
 DB 281 SerArgSerLeuThrIleSerIleValAlaAlaAlaGlyArgGluLeuPheMetThrAsp 300
 QY 997 CGGAGCGGCTGCGACAGCGCGGCGAGCTGAGCTGCAGCGCAGGACCTTCTCATGACG 1056
 DB 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuMetGln 320
 QY 1057 AAGCGCTGGCGATGGAGTCCCAACAGATCTCCAGGAGCAGCAGAGATGGAGCGGCAA 1116
 DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln 340
 QY 1117 AGGAGAAAGAAATTTGCCAGAGGACGAGCAGAGAAATAGAGATACCGGAGGAGATG 1176
 DB 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
 QY 1177 GAACAGATTGTAGAGGAGGAGAGAGTATTAGAGCAATGGAGAGCACTGGGGCTCA 1236
 DB 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380
 QY 1237 AAGGAACAGCTACTCTTGCTAAAACCATCTACTGTGAGGTACACCCAGTACCCCTTGC 1296
 DB 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 QY 1297 AAGCCAAAGATGATCAGGGAGTGAACCTGAGCTGAGCCCGCAGATGACCTGGATGGA 1356
 DB 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
 QY 1357 GGCAGGAGGAGGAGGAGGAGGAGG----- 1381
 DB 421 GlyThrGluGluGluGlyGlu-GlnProGlnGluMetLeuLysArgMetValTyrG 440
 QY 1382 -----ATTCCGGAATATAGGAGGCTTTGACCCCTACTCTATG 1422
 DB 440 nAspSerIleGlnAspLysIleSerGlyAsnMetArgLysAlaLeuThrProThrLeuCy 460
 QY 1423 TTCACCCAGACGACATCATGGGAGGATGTCGGCTCTCAGCATCAAGAAGGAGGA 1482
 DB 460 sSerProGlnSerArgSerTrpGlyArgMetSerGlySerTyrAlaSerArgArgAs 480
 QY 1483 TCCT 1486
 DB 480 pPro 481
 RESULT 4
 ID AAY07099 standard; Protein; 521 AA.
 AC AAY07099;
 XX 02-JUL-1999 (first entry)
 DE Colon cancer associated antigen precursor sequence.
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.
 OS Homo sapiens.
 XX WO9904265-A2.
 PN 28-JAN-1999.
 XX 15-JUL-1998; 98WO-US14679.
 PR 22-JUN-1998; 98US-0102322.
 PR 17-JUL-1997; 97US-0896164.
 PR 10-OCT-1997; 97US-0061599.
 PR 10-OCT-1997; 97US-0061765.
 PR 10-OCT-1997; 97US-0948705.
 PR 11-OCT-1997; 97GB-0021697.
 XX

PA (LUDW-) LUDWIG INST CANCER RES.
 XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 PI Pfeundschoh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 XX WPI: 1999-132448/11.
 XX New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 XX
 XX Disclosure: Page 666-667; 787pp; English.
 XX
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 XX
 SQ Sequence 521 AA;
 Alignment Scores:
 Pred. No.: 1.23e-186 Length: 521
 Score: 2311.00 Matches: 462
 Percent Similarity: 88.70% Conservative: 1
 Best Local Similarity: 88.51% Mismatches: 0
 Query Match: 56.09% Indels: 59
 DB: 20 Gaps: 1
 US-09-502-945-4 (1-2236) x AAY07099 (1-521)
 QY 97 ATGGACCGAAAGTGCCCGAGAAATCCCGCATAGGTGGATTCTTGATTAATAATGAT 156
 DB 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
 QY 157 GCAGAGAAGGACTATCTCTATGATGTCGCGAATGTACCACAGACCATGACGTGGCC 216
 DB 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
 QY 217 GTGCTCGTGGGACCTGAAGCTGGTCATCAATGAACCCAGCGCTGCTCTGTGTTGAT 276
 DB 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
 QY 277 GCATTCGCGCGCTGATCCCACTGAGCACCAGGTGGATATGATCAGCTGACCCCGG 336
 DB 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 QY 337 CGCTCCAGGAAGCTGAAGGAGTGCCTCTGACCGCTCTGCACCCGAGGCGCTCGGCGCTG 396
 DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
 QY 397 AGTGTGCGGTGGCTGGAGTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGGC 456
 DB 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
 QY 457 GGTCAAGCAGACAGCTCGGCTCCAGGTAGGGGAGAGATCTCCGGATCATGATAT 516
 DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 QY 517 TCCATCTCCCTCTGATCCCATGAGGAGTTCATCAACTTCATTCGAAACCAAGAACTGTG 576
 DB 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleLeuThrLysThrVal 160

QY 577 TCCATCAAGTAGAGACACATCGGCTGATCCCGTGAARAGCTCTCTGATGAGCCCTC 636
 DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
 QY 637 ACTTGGCAGTATGTGGATCAGTTGTCTCGGAATCTCGGGCGCTGGAGGACGCTGGC 696
 DB 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200
 QY 697 TCCCTCGAAATCGGAAAAACAAGGAGAAGAGTCTTCATCAGCCTGTAGGCTCCCGA 756
 DB 201 SerProGlyAsnArgGluAsnLysLysValPheIleSerLeuValGlySerArg 220
 QY 757 GGCCTTGGCTGCAGCATTTCCAGCGGCCCATCCAGAGCCTGCCATCTTTATCAGCCAT 816
 DB 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
 QY 817 GTGAAACTGGCTCCCTCTGCTGAGTGGGATGGAGATAGGGACACAGATTGTGCAA 876
 DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
 QY 877 GTCAATGGCTGCTGACTTCTTAACCTGGATCACAAGAGGCTGTAAATGTGTAATAAT 936
 DB 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
 QY 937 AGCCGCGACCTGACCATCTCCATTGTAGTGCAGCTGGCGGGAGCTGTTTCATCAGAC 996
 DB 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
 QY 997 CGGAGGCGCTGGCAGAGCGCGCAGCTGAGCTGCAGCGGCGAGGAGCTTCTCATGCAG 1056
 DB 301 ArgGluArgLeuAlaGluAlaArgGluArgGluLeuGlnArgGlnLeuLeuMetGln 320
 QY 1057 AAGCGCTGGCGATGGATGCCAACAGATCTCCAGAGCAGCAGGAGATGGAGCGCAA 1116
 DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln 340
 QY 1117 AGGAAAAGAAATGCCCCAGAGCAGAGAGAAATGAGAGATACCGGAGAGAGATG 1176
 DB 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluLeuAsnGluArgTyrArgLysGluMet 360
 QY 1177 GAACAGATTGTAGAGGAGGAAGAAGTTTAAAGAACAATGGGAAGAAGATGGGGGTCA 1236
 DB 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluLysTrpGlySer 380
 QY 1237 AAGCAACAGCTACTCTTTCCTTAAACCATCATCTCTGAGGTACACCCAGTACCCCTGCG 1296
 DB 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 QY 1297 AAGCAAAGTATGATCAGGAGTGAACCTGAGCTCGAGCCCGCAGATGACCTGATGGA 1356
 DB 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
 QY 1357 GGCAGGAGGAGCAGGAGAGACAGCAGG----- 1381
 DB 421 GlyThrGluGluGlnGlyGlu-GlnThrPheCysProSerProGlnProProArgGlyPr 440
 QY 1381 ----- 1381
 DB 440 GglyValSerThrIleSerLysProValMetValHisGlnGluProAsnPheIleTyrAr 460
 QY 1381 ----- 1381
 DB 460 gProAlaValLysSerGluValLeuProGlnGluMetLeuLysArgMetValValTyrGl 480
 QY 1382 -----ATTTCGGAAATATAGGAAGGCTTTGACCCCTACTCTATG 1422
 DB 480 nAspSerIleGlnAspLysIleSerGlyAsnMetArgLysAlaLeuThrProThrLeuCy 500
 QY 1423 TTCACCCAGCAGACATCATGGGAGGATGTCGGCTCTACCGCATCAAGAGAGGAGA 1482
 DB 500 sSerProGlnSerArgSerTrpGlyArgMetSerGlySerTyrAlaSerArgArgAs 520
 QY 1483 TCCT 1486

Db	520	pPro 521		QY	97	ATGACCCGAAAGTGGCCCGGAGAAATTCGGCATAGGTGGATTTCTGATTGAAATGAT	156
Db				Db	1	MetAspArgLysValAlaAargGluPheArgHisLysValAspPheLeuLeuGluAsnAsp	20
RESULT 5				QY	157	GCAGAGAAGGACTATCTCTATGATGCTGCGAAATGATACCAAGAGACCATGAGTGGCC	216
ID	AA07095	standard; Protein; 403 AA.		Db	21	AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla	40
XX	AC	AA07095;		QY	217	GTGCTCGTGGAGACCTGAAAGCTGGTCAATCAATGAACCCAGCCGCTGCTCTTTGAT	276
XX	DT	02-JUL-1999 (first entry)		Db	41	ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp	60
XX	DE	Colon cancer associated antigen precursor sequence.		QY	277	GCCATTGGCCGCTGATCCCACTGAAGCACCAGGTGAATATGATGACCTGACCCCGG	336
XX	KW	Cancer associated antigen; diagnosis; research; treatment; human;		Db	61	AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg	80
XX	KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;		QY	337	CGCTCCAGGAAGCTGAAGGAGGTGCTGACCCGCTGACCCCGGAGGCTCGGCGTG	396
XX	OS	Homo sapiens.		Db	81	ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu	100
XX	PN	WO9904265-A2.		QY	397	AGTGTGCTGCTGGCTGGAGTTTGGCTGGGCTCTTCATCTCCCACTCATCAAAAGC	456
XX	PD	28-JAN-1999.		Db	101	SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuLeLysGly	120
XX	PF	15-JUL-1998; 98WO-US14679.		QY	457	GCTCAGCAGACAGCTCGGCTCCAGGTAGGAGGACAGATCGTCCGGATCAATGGATAT	516
XX	PR	22-JUN-1998; 98US-0102322.		Db	121	GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr	140
XX	PR	17-JUL-1997; 97US-0896164.		QY	517	TCCATCTCCTCCTGTACCCATGAGGAGGTCTCAACCTCATTCGAACCAAGAAACCTGTG	576
XX	PR	10-OCT-1997; 97US-0061599.		Db	141	SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal	160
XX	PR	10-OCT-1997; 97US-0061765.		QY	577	TCCATCAAGTAGAGACACATCGGCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTC	636
XX	PR	10-OCT-1997; 97US-0948705.		Db	161	SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu	180
XX	PR	11-OCT-1997; 97GB-0021697.		QY	637	ACTTGGCAGTATGAGTATGCTGCTGCGAATCTGCGAATCTGGGCGCTGGAGCCGCTGGGC	696
XX	PA	(LUDW-) LUDWIG INST CANCER RES.		Db	181	ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlySerLeuGly	200
XX	PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;		QY	697	TCCCTCGAAATCGGGAACAAAGAGAGAGAGTCTTTCATCAGCCCTGGTAGCTCCCGA	756
XX	PI	pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;		Db	201	SerProGlyAsnArgGluAsnLysGluLysValPheIleSerLeuValGlySerArg	220
XX	PI	Tureci O;		QY	757	GGCTTGGCTGACGATTTCCAGCGGCCCATCCAGAGCCTGGCATCTTTATCAGCCAT	816
XX	DR	WPI; 1999-132448/11.		Db	221	GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis	240
XX	PT	New isolated cancer associated nucleic acids and polypeptides -		QY	817	GTAAACCTGGCTCCCTGCTGCTGAGTGGGATGGAGATAGGGACAGATGTCGAA	876
XX	PT	isolated using sera from cancer patients, used to develop products		Db	241	ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu	260
XX	PT	for the diagnosis, monitoring or treatment of cancers		QY	877	GTCAATGGCTGACCTTCTTAACCTGGATCAAGAGGCTGTAATGTGCTGAAATAT	936
XX	PS	Disclosure; Page 659-660; 787pp; English.		Db	261	ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn	280
XX	CC	The invention relates to a method for diagnosing a disorder characterised		QY	937	AGCCGAGCCTGACCATCTCCATTTGATGTCAGCTGGCGGAGCTGTTTCATGACAGAC	996
XX	CC	by expression of a human cancer associated antigen precursor coded for by		Db	281	SerArgSerLeuThrIleSerIleValAlaAlaAlaGlyArgGluLeuPheMetThrAsp	300
XX	CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a		QY	997	CGGAGCGGCTGGCAGCGGCGGAGCTGAGCTGCAGCGGAGGAGCTTCTCATGACAG	1056
XX	CC	biological sample isolated from a subject with an agent that specifically		Db	301	ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuLeuMetGln	320
XX	CC	binds to the NAM, an expression product or a fragment of an expression		QY	1057	AAGCGGCTGGGATGAGTCCACACAGATCCCTCCAGGACGACGAGAGATGGAGCGGCAA	1116
XX	CC	product complexed with an HLA molecule; and (b) determining the		Db	321	LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln	340
XX	CC	interaction between the agent and the NAM or the expression product as a		QY	1117	AGGAGAAAGAAATTTGCCAGAGGACGACGAGAGAAATAGAGATACCCGAGAGGATG	1176
XX	CC	determination of the disorder. The products and methods can be used in		Db	341	ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet	360
XX	CC	the diagnosis, monitoring, research, or treatment of conditions		QY	1177	GAACAGATTGTAGAGGAGGAGAGAGAGATTAAAGAACCAATGGGAAGAGACTGGGCTCA	1236
XX	CC	characterised by the expression of various cancer associated antigens.					
XX	CC	The invention provides nucleic acid sequences and encoded polypeptides					
XX	CC	which are cancer associated antigen precursors expressed in human breast					
XX	CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and					
XX	CC	lung cancer.					

SQ Sequence 403 AA;

Alignment Scores:

Pred. No.:	2,9e-164	Length:	403
Score:	2046.00	Matches:	403
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	49.66%	Indels:	0
DB:	20	Gaps:	0

US-09-502-945-4 (1-2236) x AA07095 (1-403)

|||||
Db 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380
QY 1237 AAGGAACAGCTACTTGGCTAAACCATCTGCTGAGGTACACCCAGTACCCCTTGGC 1296
|||||
Db 381 LysGluGlnLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCCAAAG 1305
|||||
Db 401 LysProLys 403

RESULT 6

AA07097
ID AAY07097 standard; Protein; 328 AA.

XX
AC AAY07097;

XX
DT 02-JUL-1999 (first entry)

XX
DE Colon cancer associated antigen precursor sequence.

XX
KW Cancer associated antigen; diagnosis; research; human; lung cancer;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.

XX
OS Homo sapiens.

XX
PN W09504265-A2.

XX
PD 28-JAN-1999.

XX
PF 15-JUL-1998; 98WO-US14679.

XX
PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

XX
(LUDW-) LUDWIG INST CANCER RES.

XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;

PI Tureci O;

XX
WPI; 1999-132448/11.

XX
PS New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products

PT for the diagnosis, monitoring or treatment of cancers

XX
Disclosure; Page 662-663; 787pp; English.

XX
CC The invention relates to a method for diagnosing a disorder characterised

CC by expression of a human cancer associated antigen precursor coded for by

CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a

CC biological sample isolated from a subject with an agent that specifically

CC binds to the NAM, an expression product or a fragment of an expression

CC product complexed with an HLA molecule; and (b) determining the

CC interaction between the agent and the NAM or the expression product as a

CC determination of the disorder. The products and methods can be used in

CC the diagnosis, monitoring, research, or treatment of conditions

CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides

CC which are cancer associated antigen precursors expressed in human breast

CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

XX lung cancer.

Score: 1732.00 Matches: 328
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.04% Indels: 0
DB: 20 Gaps: 0

US-09-502-945-4 (1-2236) x AAY07097 (1-328)

QY 1069 ATGGAGTCCCAACAGATCTCCAGAGCAGCAGAGTGGAGCGCAAGAGAGAAAGAA 1128
|||||
Db 1 MetGluSerAsnLysIleLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 20

QY 1129 ATTGCCAAGAAGCAGAGAGAGAAATGAGAGATACCGGAAGAGATGGAACATTGTA 1188
|||||
Db 21 IleAlaGlnLysAlaAlaGluLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40

QY 1189 GAGGAGGAAGAGAGTTTAAGAACATGGGAAGAAAGACTGGGGCTCAAGAGACAGCTA 1248
|||||
Db 41 GluGluGluGluLysPheLysLysGlnTrpGluLysTrpGlySerLysGluGlnLeu 60

QY 1249 CTCTTGGCTAAACCATCTGAGTACACCCAGTACCCCTTTCGCAAGCCAAAGTAT 1308
|||||
Db 61 LeuLeuProLysThrIleThrAlaGluValHisProValProLeuArgLysProLysTyr 80

QY 1309 GATCAGGAGTGGAACTGAGCTCGAGCCCGCAGATGACCTGGATGGAGGACGAGGAG 1368
|||||
Db 81 AspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGlyGlyThrGluGlu 100

QY 1369 CAGGAGAGCAGGATTTCCGGAAATATGAGGAGGCTTTGACCCCTACTCTATGTTACC 1428
|||||
Db 101 GlnGlyGluGlnAspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThr 120

QY 1429 CCAGAGCAGATCATGGGAAGGATGTCGGCTCTACCATCAAGAGGAGGATCTCTTA 1488
|||||
Db 121 ProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeu 140

QY 1489 GACTTGGCCCTGGAAGCGGTGTGAGCTCCCCATTTGGGAAGGTGGTCTGCTCTGTG 1548
|||||
Db 141 AspLeuAlaLeuGluGlyGlyValAspSerProIleGlyLysValValValSerAlaVal 160

QY 1549 TATGAGCGGGGAGCTGCTGAGCGGCATGTGGCATGTGGAAGGGGAGGATCATGGCA 1608
|||||
Db 161 TyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAla 180

QY 1609 ATCAACGGCAAGATTGTACAGACTACACCTCGGTGAGGCTGAGCTGCGCTCGCAGAG 1668
|||||
Db 181 IleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAlaLeuGlnLys 200

QY 1669 GCCTGGAATCAGGCGGGGAGTGGATCGACCTTGTGGTGGTGGTGGTGGTGGTGGTGG 1728
|||||
Db 201 AlaTrpAsnGlnGlyAspTrpIleAspLeuValAlaValAlaValCysProProLysGlu 220

QY 1729 TATGACGATGAGCTGACCTTCTGCTGAAGTCCAAAGGGGAAACCAATTCACCGCTTA 1788
|||||
Db 221 TyrAspAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnGlnIleHisAlaLeu 240

QY 1789 GGAACAGTGAAGTCTCGGCGCCACCTCGTGAACACAAAGCCCTCGGACAGCCTTGAGAGA 1848
|||||
Db 241 GlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgThrSerLeuGluArg 260

QY 1849 GGCCACATGACACACACAGATGGATCGTCTGGGACCTGAATCTATCACCAGGATCTC 1908
|||||
Db 261 GlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSerProArgAsnLeu 280

QY 1909 AAATCCCTTTGGCCCTCAACAGCGCCAGATAAGGAACAGCTCGGCGCACCTTTTGA 1968
|||||
Db 281 LysLeuProLeuAlaLeuAsnGlnGlnIleArgAsnSerSerGlyHisPheGlu 300

QY 1969 GGCCCAATGTGGAGGAAGGAGCAGCCAGCCGCTTTGGGAGAGATCTCAAGSATCCAGAC 2028
|||||
Db 301 GlyGlnCysGlyGlyGlyAlaAlaSerArgLeuGluAspLeuLysAspProAsp 320

QY 2029 TCTCATCTCTTCTCTGCGCCAG 2052

Alignment Scores:

Pred. No.: 9.79e-138 Length: 328

KW infectious disease; cardiovascular disorder.
XX Homo sapiens.
XX WO200055351-A1.
XX PD 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US05883.
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX WPI; 2000-587534/55.
XX DR N-PSDB; AAC97995.
XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
XX Claim 11; Page 1339; 2104pp; English.
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
XX invention.
XX
XX SQ Sequence 120 AA;
Alignment Scores:
Pred. No.: 3.33e-29 Length: 120
Score: 447.00 Matches: 92
Percent Similarity: 87.04% Conservative: 2
Best Local Similarity: 85.19% Mismatches: 4
Query Match: 10.85% Indels: 10
DB: 21 Gaps: 2
US-09-502-945-4 (1-2236) x AAB53238 (1-120)
QY 1315 GGAGTGAACCTGAGCTCGAGCGCGCAGATGACCTGGATGAGGACGAGGAGGAGG 1374
Db 23 GlyLeuArgProValLeuGlnPro-----ArgGlnGly 33
QY 1375 GAGCAGGATTCGCGAAATATGAGGAGGCTTTGACCCCTACTCTATGTTACCCCGAG 1434
Db 34 ---GlnAspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGlu 52
QY 1435 CAGATCATGGGGAAGGATGTCCGCTCTACGCATCAAGAGGAGGAGGATCCTTTAGACCTG 1494
Db 53 GlnIleMetGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeuAspLeu 72
QY 1495 GCCTTGAAGCGGTGTGGACTCCCATTTGGAGGTGTCTTCTGCTGTGTATGAG 1554
Db 73 AlaLeuGluGlyGlyValAspSer***IleGlyLysValValValSerAlaValTyrGlu 92
QY 1555 CGGGGACCTGCTGAGCGGCATGTGGCATTTGAAAGGGACGAGATCATGGCAATCAAC 1614
Db 93 ArgGlyAlaAlaGluArgHisGlyGlyValValValGlyAspGluIleMetAlaIleAsn 112

QY 1615 GGCAAGATTGTGACAGACTACACC 1638
Db 113 GlyLysIleValThrAspTyrThr 120
RESULT 9
ABB61924
ID ABB61924 standard; Protein; 493 AA.
XX AC ABB61924;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 12564.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX OS
XX PN WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL06027.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 12564; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 493 AA;
Alignment Scores:
Pred. No.: 4.24e-27 Length: 493
Score: 425.50 Matches: 128
Percent Similarity: 42.76% Conservative: 58
Best Local Similarity: 29.43% Mismatches: 164
Query Match: 10.33% Indels: 85
DB: 22 Gaps: 11
US-09-502-945-4 (1-2236) x ABB61924 (1-493)
QY 340 TCCAGAGACTGAAGAGGTGGTGTGACCGCTGTGCACCCGAGAGGCTC----- 390
Db 58 SerSerArgLeuArgValLeuArgLeuValArgProHisGlnArgArgLeuValGly 77
QY 391 -----GSCCTGAGTGTGCGTGGCTGGCTGAGTTTGGCTGT 426
Db 78 GlyProGluArgGlySerThrTyrGlyPheThrValArgGlyGlyArgGluHisGlyThr 97

Qy	427	GGGCTCTTCATCTCCACCTCATCAAAAGCGGTGCAGGCAGACAGCGTCCGGCTCCAGGTA	480
Db	98	GlyPhePheValSerHisValGluHisGlyGlyGluAlaHisLeuLysGlyLeuArgIle	117
Qy	487	GGGACGAGATCGTCGGATCAATGGATATTCCATCTCCTGTACCACCTCAGGAGGTC	546
Db	118	GlyAspGlnIleLeuArgIleAsnGlyPheArgLeuAspAlaValHisLysGluPhe	137
Qy	547	ATCAACCTCATTCGAACCAAGAATACTGTCCTCATCAAGTGAGACATCGGCTGATC	606
Db	138	IleGlnLeuValAlaGlyGlnAspArgValThrLeuLysValArgGlyValGlyMetLeu	157
Qy	607	CCCCGAAAAGCTCTCTGATGAGCCCCTCACTTGCCAGTATGTGGATCAGTTTGTGTCG	666
Db	158	ProValargAspLeuProGluGluArgLeuSerTrpSerValValLys-----LeuPro	175
Qy	667	GAATCTGGGCGTGGAGCGACGCTGGGCTCCCTCGAAATCGGAAAAACAAGAGAGAAG	726
Db	176	SerValSerGlyThrProSerGluSerSerPheLysGlyGluArgArgGlyAlaSerArg	195
Qy	727	AAGTCCTTCATCAGCCTG---GTAGCTCCCGA---GGCCTTGGCTGCAGCATTTCCACG	780
Db	196	AspIleSerValValLeuHisValAlaProArgThrLysLeuGlyLeuGlyIleCysLys	215
Qy	781	GGCCCCATCCAGAAGCTGCATCTTTATCAGCCATGTCAAACCTGGCTCCGCTCTGCT	840
Db	216	GlyProGluTrpLysProGlyIlePheValGlnPheThrLysAspArgSerValAlaArg	235
Qy	841	GAGTGGGATGGAGATAGGGGACCAGATTGTCCAATCGAATGGGCTGCATCTCTCAAC	900
Db	236	GluAlaGlyLeuArgProGlyAspGlnIleLeuSerValAsnSerIleaspPheSerAsp	255
Qy	901	CTGGATCACAGGAGGCTGAATGTGCTGAAAAATAGCCGCGAGCTGACCATCTCCATT	960
Db	256	ValLeuPheSerGluAlaValAlaValMetLysSerSerSerLysLeuAspMetValVal	275
Qy	961	GTAGCTCAGCTGCGCGGAGCTGTTCATGACAGACGCGGAGCGCTGSCAGAGCGCGG	1020
Db	276	ArgThrAlaAlaGlyCysAspLeuPheProGlyGlu-SerSerGlyTyrrAsnSerSerAl	295
Qy	1021	CAGGTCAGCTGACGGCGGAGGACTTCTCATGCAGAACGGCTCGCATGGAGTCCCAAC	1080
Db	295	aSer--SerValThrGlyAspGlnSerProCys-----TpAlaAspAl	309
Qy	1081	AAGATCTCCACGACGACGAGGAGATGGAGCGCAAGGAGAAAGAAATTCGCCAGAG	1140
Db	309	aLysSerLysArg-LeuThr-----AlaValArgGluGluSerGlyAlaGlyGlyG	326
Qy	1141	GCAGCAGAGGAAATGAGATACCCGGAAGGAGATGGACAGATTGTAGAGGAGGAAG	1200
Db	326	lyGlyCysGlyLeuSerSerAlaProGlyAlaGlySerProAsnTrpSerGlnGlyValG	346
Qy	1201	AAGTTTAAGAAAGCAATGGGAGAGACTG-----	1229
Db	346	luValHisLysGlnMetAsnLysThrIleIleLysLeuThrgluAsnGlyThrSerIleA	366
Qy	1230	-----GGGCTCAANGAA	1242
Db	366	snaSnrThrTyrlleAlaSerThrGlyLysSerSerValSerGlySerGlySerThrGlyS	386
Qy	1243	CAGCTACTCTTGCTTAAACCATCATCTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCA	1302
Db	386	erGlyThrSerGlyArgSerGlnGlnSerGlnSerAsnProSerAsnProSerArg---	404
Qy	1303	AAGTATGATCAGGAGTGAACCTGAGTCGAGCCCGCAGATGACCT-----	1349
Db	405	-----asnSerThrThrMetLysArgSerHisLeuArgProValAsnSerAlag	421
Qy	1350	-----GGATGGACGCGGAGGAGGAGGACGAGGATTTCGGGAATATGAG	1398
Db	421	lySerGlylleGlyLeuSerSerGlySerAlaglySerAlacly-----	435

QY	1399	GAAGCCTTTGACCCCTACTCTATGTTCA	CCCCAGACAGATCATGGGGAAGGATGTCGGG	1458
Db	436	-----	-----SerAlaGlySerSerG	441
QY	1459	CTCTACGCATCAAGAAGGAGGAGGATCCTTAGACCTGGCCCT	1499	
Db	441	lySerGlySerArgSerGlyGlyValIleAlaProAlaPro	454	
RESULT 10				
ID	AY38390			
XX	AY38390 standard; Protein; 98 AA.			
AC	AY38390;			
XX				
DT	30-SEP-1999 (first entry)			
XX				
DE	Human secreted protein encoded by gene No. 5.			
XX				
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;			
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;			
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;			
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;			
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;			
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;			
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;			
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.			
OS	Homo sapiens.			
XX				
PN	W09935158-A1.			
XX				
PD	15-JUL-1999.			
XX				
PF	06-JAN-1999; 99WO-US00108.			
XX				
PR	07-JAN-1998; 98US-0070704.			
PR	07-JAN-1998; 98US-0070657.			
PR	07-JAN-1998; 98US-0070658.			
PR	07-JAN-1998; 98US-0070692.			
XX				
PA	(HUMA-) HUMAN GENOME SCI INC.			
XX				
PI	Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;			
PI	Olsen HS, Rosen CA, Ruben SM, Soppet DR;			
XX				
DR	WPI; 1999-444190/37.			
DR	N-PSDB; AA206223.			
XX				
PT	New isolated human genes and the secreted polypeptides they encode			
XX				
PS	Claim 11; Page 178-179; 227pp; English.			
XX				
CC	This sequence represents a secreted human protein encoded by the gene			
CC	clone detailed in the descriptor line. The gene can be used to generate			
CC	fusion proteins by linking to the gene to a human immunoglobulin Fc			
CC	portion (e.g. AA206210) for increasing the stability of the fused			
CC	protein as compared to the human protein only.			
CC	The invention relates to 36 novel genes and their fragments (nucleic			
CC	acid sequences: AA206219-206263; amino acid sequences AA38386-Y38498)			
CC	which are useful for preventing, treating or ameliorating medical			
CC	conditions e.g. by protein or gene therapy. Also, pathological			
CC	conditions can be diagnosed by determining the amount of the new			
CC	polypeptides in a sample or by determining the presence of mutations in			
CC	the new polynucleotides. Specific uses are described for each of the 36			
CC	polynucleotides, based on which tissues they are most highly expressed in			
CC	(see AA206219 for described uses).			
XX				
SQ	Sequence 98 AA;			
Alignment Scores:				
Pred. No.:	1.05e-26	Length:	98	
Score:	417.00	Matches:	88	
Percent Similarity:	83.81%	Conservative:	0	

Alignment Scores:		
Pred. No.:	1.05e-26	98
Score:	417.00	88
Percent Similarity:	83.81%	0
Length:		98
Matches:		88
Conservative:		0

Best Local Similarity: 83.81% Mismatches: 0
Query Match: 10.12% Indels: 17
DB: 20 Gaps: 1

US-09-502-945-4 (1-2236) x AAY38390 (1-98)

QY 676 GCGTCGAGGAGCCTGGCTCCCTCGAATCGGGAACAGAGGAGAAAGTCTTC 735
DB 10 GlyValArgGlySerLeuGlySerProGlyAsnArgGluAsnLysGluLysValPhe 29

QY 736 ATCAACCTGGTGGTCCCGAGGCTTGGCTGCAGCATTCAGCGCCCTCCAGAG 795
DB 30 IleSerLeuValGlySerArgGlyLeuGlyCysSerIleSerGlyProIleGlnLys 49

QY 796 CTGGCATCTTTATCAGCATGTGAACCTGGCTCCCTGCTCTGCTGAGGTGGATTGG 855
DB 50 ProGlyIlePheIleSerHisValLysProGlySerLeuSerAlaGluValGlyLeuGlu 69

QY 856 ATAGGGACAGATTGTGAAGTCAATGGCTGCGACTTCTTAACCTGGATCACAAGAG 915
DB 70 IleGlyAspGlnIleValGluValAsnGlyValAspPheSerAsnLeuAspHisLysGlu 89

QY 916 GCTGTAATGTCTGAAATAGCCGACCTGACCATCTCCATTGTAGCTGCGCTGCG 975
DB 90 -Leu-----GlnLeuAl 93

QY 976 CGGAGCTGTTC 988
DB 93 aGlySerCysSer 97

RESULT 11
AAY38437
ID AAY38437 standard; Protein; 95 AA.
AC AAY38437;
XX
DT 18-OCT-1999 (first entry)
XX
DE Human secreted protein encoded by gene No. 5.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO935158-A1.
XX
PD 15-JUL-1999.
XX
PF 06-JAN-1999; 99WO-US00108.
XX
PR 07-JAN-1998; 98US-0070704.
XX
PR 07-JAN-1998; 98US-0070657.
XX
PR 07-JAN-1998; 98US-0070658.
XX
PR 07-JAN-1998; 98US-0070692.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
XX
DR WPI; 1999-444190/37.
DR N-PSDB; AA206223.
XX
PT New isolated human genes and the secreted polypeptides they encode
XX Disclosure; Page 205; 227pp; English.

XX This sequence represents a secreted human protein encoded by the gene
CC clone detailed in the descriptor line. The gene can be used to generate
CC fusion proteins by linking to the gene to a human immunoglobulin Fc
CC portion (e.g. AAZ06210) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 36 novel genes and their fragments (nucleic
CC acid sequences: AA06219-206263; amino acid sequences: AAY38386-Y38498)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 36
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAZ06219 for described uses).

XX
SQ Sequence 95 AA:

Alignment Scores: 8.81e-26 Length: 95
Pred. No.: 406.00 Matches: 90
Score: 60.13% Conservative: 2
Percent Similarity: 58.82% Mismatches: 1
Best Local Similarity: 9.85% Indels: 60
Query Match: 20 Gaps: 1
DB: 1

US-09-502-945-4 (1-2236) x AAY38437 (1-95)

QY 1752 GCTGAAGTCCAAAGGGGAAACCAATTCACGCGTTAGGAACAGTCCGCGCCCA 1811
DB 3 SerGluValGlnLysGlyLysProAsnSer-ProLeuGlyAsnSerGluLeuArgProH1 22

QY 1812 CCTGTGAACACAAAGCCTCGACAGCCTTGAGAGAGGCCACATGACACACACAGATG 1871
DB 22 sLeuValAsnThrLysProArgThrSerLeuGluArgGlyH1s ----- 36

QY 1872 GCATCTTGGGACCTGAATCTATCACCAGGAAATCTCAAACTCCCTTTGGCCTGAACCA 1931
DB 36 ----- 36

QY 1932 GGGCCAGATAAGGAACAGCTCGGGCCACTTTTGAAGGCCAATGTGAGGAAGGGAGC 1991
DB 36 ----- 36

QY 1992 AGCCAGCCGTTGGGAGAAGATCTCAAGATGCCAGACTCTCATCTTCTCTGCGCCA 2051
DB 37 -----Thr-IleProPheLeuTrpProS 44

QY 2052 GTGAATTTGGTCTCTCCAGCTTTGGGGGACTCTTCCCTTGAACCTTAATAGACCCAC 2111
DB 44 erGluPheGlyLeuSerGlnLeuTrpGlyThrProSerLeuAsnProAsnLysThrProL 64

QY 2112 TGGAGTCTCTCTCTCCATCCCTCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2171
DB 64 euGluSerLeuSerLeuHisProSerProLeuProSerAlaLeuIleAlaAlaArgIleV 84

QY 2172 TCACCTCCAAACCTTACTCTGAGCTCATTAATAAAA 2206
DB 84 alThrProAsnLeuThrLeuSerSerLeuIleLys 95

RESULT 12
ABG22366
ID ABG22366 standard; Protein; 1037 AA.
XX
AC ABG22366;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22357.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX

QY 1224 AGACTGGGGCTCAAGGACACCTACTCTTGCCTAAACCACTCACTGCTGAGGTACACCC 1283
Db 467 -----LeuProArgGluSerSerTyrAspIleTyrAr 477
QY 1284 ACTACCCCTTCGCAAGCAAAATGATGACAGGAGCTGGAACCTGAGCTCGAGCCGCAGA 1343
Db 477 gValProSerSerGlnSerMetGluAspArgGlyTyrSerPro----- 491
QY 1344 TGACCTGGATGAGGACGAGGAGGAGGAGGAGGAGGATTCCTCGGAATATGAGGAAGG 1403
Db 491 ----- 491
QY 1404 CTTTGACCCCTACTATGTTTCAACCCAGAGCAGATCATGCGGAAGGATGCGCGCTCCT 1463
Db 492 -----AspThrArgValVa 496
QY 1464 ACGCATCAGAGAGGAGGATCTTACACCTGCGCCCTGGAGCGCTGTGGACTCCCGCAT 1523
Db 496 lArgPheLeuLysGlyLysSerIleGlyLeuArgLeuAlaGlyGlyAsnAsp-----Va 514
QY 1524 TGGGAAGTGGCTGCTGCTGCTATGAGCGGGGAGCTGTGAGCGGCATGTTGGCAT 1583
Db 514 lGly---IlePheValSerGlyVal---GlnAlaGlySerProAlaAspGlyGlnGlyI 532
QY 1584 TGTAAAGGGAGGAGATCATGCGCAATCAACGCGAAGATTGACAGACTACACCTGGC 1643
Db 532 eGlnGluGlyAspGlnIleLeuGlnValAsnAspValProPheGlnAsnLeuThrArgGl 552
QY 1644 TGAGGCTGACGCTGCGCTGCAGAGGCGCTGGAATCAGGCGGGGACTGGATCGACCTTGT 1703
Db 552 uGluAla-----ValGlnPheLe 558
QY 1704 GGTTCGCGCTCGCCCGCCAAAGGAGTATGAC-----GATGAGCT 1742
Db 558 uLeuGlyLeuProProGlyGluGluMetGluLeuValThrGlnArgLysGlnAspIlePh 578
QY 1743 GACCTTCCTGCTGAAGTCAAAAGGGGAAACCAATTCACGGCTTAGGAAACAGTGAGCT 1802
Db 578 eTrpLysMetValGlnSerArgValGlyAspSerPheTyrIleArgThrHisPheGlu 598
QY 1803 CGGGCCCGACCTCGTGAACACAAAGCGCTCGGACCGCTT-----GAGAGAGGCCACAT 1856
Db 598 uGluPro-----SerProSerGlyLeuGlyPheThrArgGlyAspVa 613
QY 1857 GACACACACC-----AGATGGCATCTTGGGACCTGTAATCATCACCCAGGAATCTCAA 1910
Db 613 lPheHisValLeuAspThrLeuHisPro----- 622
QY 1911 ACTCCCTTTGGCCCTGAACCGGCGCCAGATAGGAACACACTCGGCGCACTTTTGTGAAGG 1970
Db 623 -----GlyProGlyGlnSerHisAlaArgGlyGlyHisTrpLeu----- 635
QY 1971 CCAATGTGGAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2030
Db 636 -----AlaValArgMetGlyArgAspLeuArgGluGlnGluAr 648
QY 2031 TCATTCCTTCCT 2043
Db 648 gGlyIleIlePro 652
RESULT 13
AAE21718
ID AAE21718 standard; Protein; 928 AA.
XX
AC AAE21718;
XX
XX
DT 16-JUL-2002 (first entry)
XX
DE Human PKIN-13 protein.
XX
KW Human; kinase; enzyme; PKIN-13 protein; immune system disorder; anaemia;
KW acquired immune deficiency syndrome; thymic hypoplasia; Crohn's disease;

asthma; neurological disorder; epilepsy; Charcot-Marie-Tooth disease;
ADIS; seizures; cell proliferative disorder; cancer; adenocarcinoma;
leukaemia; lymphoma; melanoma; sarcoma; developmental disorder;
Down's syndrome; gene therapy; protein therapy; cytostatic.
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Domain 10..99
FT Domain /note= "GLGF domain"
FT Domain 20..101
FT Domain /note= "PDZ domain"
FT Domain 204..280
FT Domain /note= "PDZ domain"
FT Domain 388..469
FT Domain /note= "GLGF domain"
FT Domain 391..471
FT Domain /note= "PDZ domain"
FT Domain 429..439
FT Domain /note= "PDZ domain"
FT Domain 733..754
FT Domain /note= "Leucine zipper domain"

WO200218557-A2.

07-MAR-2002.

31-AUG-2001; 2001WO-US27219.

31-AUG-2000; 2000US-229873P.

08-SEP-2000; 2000US-231357P.

14-SEP-2000; 2000US-232654P.

22-SEP-2000; 2000US-234902P.

29-SEP-2000; 2000US-236499P.

06-OCT-2000; 2000US-238389P.

13-OCT-2000; 2000US-240542P.

(INCY-) INCYTE GENOMICS INC.

Bandman O, Nguyen DB, Walia NK, Hafalia AJA, Yao MG, Gandhi AR;

Gururajan R, Ding L, Patterson C, Yue H, Baughn MR, Tribouley CM;

Thornton M, Elliott VS, Lu Y, Ison CH, Au-Young J, Tang YT;

Azimzai Y, Burrill JD, Marcus GA, Zingler KA, Lu DAM, Lal PG;

Ramkumar J, Warren BA, Kearney L, Policky JL, Thangavelu K,

Burford N;

WPI; 2002-329769/36.

N-PSDB; AAD34310.

New human kinases, useful for diagnosing, treating or preventing immune

system disorders (e.g. Crohn's disease), neurological disorders (e.g.

epilepsy), or cell proliferative disorders (e.g. cancers such as

leukemia or lymphoma)

Claim 68; Page 174-176; 218pp; English.

The present invention relates to human kinases (PKIN) and polynucleotides

encoding such proteins. PKIN sequences of the invention are useful for

diagnosing, treating or preventing disorders associated with aberrant

expression of PKIN, particularly immune system disorders (e.g. acquired

immune deficiency syndrome (AIDS), thymic hypoplasia, Crohn's disease,

anaemia, asthma), neurological disorders (e.g. epilepsy, Charcot-Marie-

Tooth disease or seizures), cell proliferative disorders (e.g. cancers

such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma),

and developmental disorders (e.g. Down's syndrome). They are also used

in gene therapy and protein therapy. The present sequence is human

PKIN-13 protein.

XX

SQ Sequence 928 AA;

Alignment Scores: 6.11e-14 Length: 928

Pred. No.: 271.50 Matches: 142

Score:

Db	340	SerProArgLeu	-----ArgArgGluSerSerValAsp	350
QY	1135	CAGAGGCACGACGAGGAAATGAGATACCCGAGAGAGATGGACAGATTCTAGAGAG	119	
Db	351	SerArgThrIleSerGluProAspGluGlnArgSerGlu	---	
QY	1195	GAAGAGAAGTTTAAAGACCAATGGGAAGAAGACTGGGGCTCAAAGGAACAGCTACTCTTG	125	
Db	364	-----Leu	364	
QY	1255	CCTAAACCATCACTGCTGAGGTACACCCAGCATACCCCTTCGCAAGCCAAAGTATGATCAG	131	
Db	365	ProArgGluSerSeryrAspIleThrArgValProSerSerGlnSerMetGluAspArg	384	
QY	1315	GGAGTGGAACTGAGCTCGAGCCCGCAGATGACCTGATGGAGGCACGAGGACGAGGA	137	
Db	385	GlyTyrSerPro	-----388	
QY	1375	GAGCAGGATTCCGGAAATATAGGAAGGCTTTGACCCCTACTCTATGTTACCCACGAG	143	
Db	388	-----388	388	
QY	1435	CAGATCATGGGAAGGATGTCGGCTCTACGCATCAAGAGAGGGATTCCTTAGACCTG	149	
Db	389	-----AspThrArgValValArgPheLeuLysGlyLysSerIleGlyLeu	403	
QY	1495	GCCCTGAAGCGGTGTGGACTCCCCCATTTGGGAGGTGGTCTCTCTGCTGTATCATC	155	
Db	404	ArgLeuIleGlyLysAsnAsp	-----ValGly---IlePheValSerGlyVal---Gln	
QY	1555	CGGGAGCTGCTGAGCGCATGTGTCATGTGAAGGGGAGAGATCATGCGCAATCAAC	161	
Db	420	AlaGlySerProAlaaspGlyGlnGlyIleGlnGluGlyAspGlnIleLeuGlnValasn	439	
QY	1615	GGCAAGATTGTACAGACTACACCTGGCTGAGGTGAGCTGCCTGCAGAACGCCTGG	167	
Db	440	AspValProPheGlnAsnLeuThrArgGluGluAla	-----451	
QY	1675	AATCAGCGGGGACTGGATFCGACCTGTGGTGGCTGCTGCCCCCAAGAGGATATGAC	173	
Db	452	-----ValGlnPheLeuLeuGlyLeuProGlyGluGluMetGlu	465	
QY	1735	-----GATGAGCTGACCTTCTTGCTGAATCCAAAGGGGAAAC	177	
Db	466	LeuValThrGlnArgLysGlnAspIlePheTrpLysMetValGlnSerArgValGlyAsp	485	
QY	1774	CAATTCACCGCGTTAGGAACAGTGAAGTCCGGCCCGCCCTCGTGAACACAAAGCCCTCG	183	
Db	486	SerPheTrpIleArgThrHisPheGluLeuGluPro	-----SerProPro	
QY	1834	ACCAGCGCTT-----GAGAGAGCCACATGACACACAC-----AGATGGCATCCTTTGG	188	
Db	501	SerGlyLeuGlyPheThrArgGlyAspValPheHisValLeuAspThrLeuHisPro	519	
QY	1882	GACCTGAATCATCACCCAGGAATCTCAAACCTCTTTGGCCCTGCAACCGAGCGCATATA	194	
Db	520	-----GlyProGlyGlnSer	524	
QY	1942	AGGAACAGCTGGGCCCATCTTTTGAAGCCCAATGTGGAGGAAGGAGCAGCCAGCCGT	200	
Db	525	HisAlaArgglyGlyHisTrpLeu	-----AlaValArg	
QY	2002	TTGGGACAGATCTCAAGATCCAGACTCTCATTTCTTTCCT	2043	
Db	536	MetGlyArgaspLeuArgGluGlnArgGlyIleIlePro	549	
RESULT	14			
AAV53753				
ID	AAV53753	standard; Protein; 2037 AA.		
XX	AAV53753;			
XX	AAV53753;			
XX	22-FEB-2000	(first entry)		

Db 209 GlyIleSerIleValGlyGlyArgGlyMetGlySerArgLeuSerAsnGlyGluValMet 228
QY 793 AAGCTGCTGATCTTTATCAGCAGCTGAACCTGGCTCCCTGCTGCTGAGAGGTGGGA--- 849
Db 229 Arg---GlyIlePheIleHisValLeuGluAspSerProAlaGlyLysAsnGlyThr 247
QY 850 TTGGAGATAGGGGACAGATTGTGGAAGTCAATAGCGTCGACTTCTCTTAACCTGGATCAC 909
Db 248 LeuLysProGlyAspArgIleValGluValAspGlyMetAspLeuArgAlaSerHis 267
QY 910 AAGGAGGCTGTAAATGCTCTGAAATAAGCCGACGCTGACCACTCCATGTTAGCTGCA 969
Db 268 GluGlnAlaValGluAlaIleArgLysAlaGlyAsnProValPheMetValGlnSer 287
QY 970 GCTGCGGGGAGCTGTTATGACACAGCGGAGCGGCTGGCAGAGCGCGCAGCGTGAG 1029
Db 288 -----IleIleAsnArgProArgAlaProSerGlnSerGluSerGlu 301
QY 1030 CTGACGCGGAGGAGCTTCTCATGTCAGAGCGGCTGGCGATGGAGTCCAAACAAGATCCTC 1089
Db 302 ProGluLysAlaProLeuCys----- 308
QY 1090 CAGGAGCAGCAGGAGATGGAGCGGCAAGGAGAGAAATTTGCCAGAGCGCAGCAGAG 1149
Db 309 -----SerValProProProProSerAlaPheAlaGluMetGlySerAsp 324
QY 1150 GAAATATGAGATACCGGAAGAGATGGAACAGATTGTAGAGGAGGAAGAGAGTTTAAG 1209
Db 325 HisThrGlnSerSerAlaSerLysIleSerGlnAspValAspLysGluAspLupheGly 344
QY 1210 AAGCAATGGGAA-----GAAGACTGGGGCTCAAGAGAACACTACTCTTGGCTAAA 1260
Db 345 TyrSerTrpLysAsnIleArgGluArgTyrGly----- 355
QY 1261 ACCATCACTGCTGAGTACACCCAGTACCCTTCGCAAG----- 1299
Db 356 ThrLeuThrGlyGluLeuHisMetIleGluLeuGlyLysGlyHisSerGlyLeuGlyLeu 375
QY 1300 -----CCA 1302
Db 376 SerLeuAlaGlyAsnLysAspArgSerArgMetSerValPheIleValGlyIleAspPro 395
QY 1303 AAGTATGATCAGGGAGTGGAACTGAGCTCGAGCCCGCAGATGACTGGATGGAGGCAGC 1362
Db 396 AsnGlyAlaAlaGlyLysAspGlyArgLeuGlnIleAlaAspGluLeu-----Leu 412
QY 1363 GAGGAGCAGGAGCAGCAGGATTTCCGGAATATAGAGAAAGGCTTTGACCCCTACTCTATG 1422
Db 413 GluIleAsnGlyGlnIleLeuTyrGlyArgSerHisGlnAsnAlaSerSerIleIleLys 432
QY 1423 TTCACCCACAGACAGATC----- 1440
Db 433 CysAlaProSerLysValLysIleIlePheIleArgAsnLysAspAlaValAsnGlnMet 452
QY 1440 -----ATGGGGAAG 1449
Db 453 AlaValCysProGlyAsnAlaValGluProLeuProSerAsnSerAspSerLeuGlnAsn 472
QY 1441 ----- 1470
Db 473 LysGluThrGluProThrValThrThrSerAspAlaAlaValAspLeuSerSerPheLys 492
QY 1450 GATGTCGGGCTCCTACGCATC---AAGAAGAGGGATCCTTAGACCTGGCCCTGGGAAGGC 1506
Db 493 AsnValGlnHisLeuGluLeuProLysAspGlnGlyGlyLeuGlyIleAlaIle----- 510
QY 1507 GGTGTGGACTCCCCCATTTGGGAGGTGTCGTTCTGCTGTATGAGCGGGGAGCTGCT 1566
Db 511 SerGluGluAspThrLeuSerGlyIleIleLysSerLeuThrGluHisGlyValAla 530
QY 1567 GACCGCATGGTGGCATTTGTAAAGGGGAGCAGATCATCGCAATCAACGCAAGATTGTG 1626
Db 531 AlaThrAspGlyArgLeuLysValGlyAspGlnIleLeuAlaValAspAspGluIleVal 550

QY 1627 ACAGACTACACCTGGCTGAGCTGACGCTGCCCTGCAGGAAGGCCTCGAATCAGGCGGG 1686
Db 551 ValGlyTyrProIleGluLysPheIleSerLeuLeuLysThrAla----- 565
QY 1687 GACTGGATCAGCTTGTGGTGGCTGCTGCCCCCAAGAGTATGACGATGAGCTGACC 1746
Db 566 -----LysMetThr 568
QY 1747 TTCTTGTCTGAA-GTCCAAAAGGGGAAACCAAAATTCACGCGTTAGGAAACAGTGAAGTCCG 1805
Db 569 ValLysLeuThrIleHisAlaGluAsnProAspSerGlnAlaValProSerAlaAlaGly 588
QY 1806 GCCCCACCTCGTGAACACAAAGCCTCGACAGCCTTGAGAGAGGCCACATGACACACAC 1865
Db 589 AlaAlaSerGlyGluLysLysAsnSerSerGlnSerLeuMetValPro----- 604
QY 1866 CAGATGCGCATCCTGGGACCTGAATCTATC-----ACCCAGGAATCTCAAACTCCC- 1916
Db 605 GlnSerGlySerProGluProGluSerIleArgAsnThrSerArgSerThrProAla 624
QY 1917 ---TTTGGCCCTGAACCA----- 1931
Db 625 IlePheAlaSerAspProAlaThrCysProIleIleProGlyCysGluThrThrIleGlu 644
QY 1932 -----GGGCCAGATAAGGAACAGCTC 1952
Db 645 IleSerLysGlyArgThrGlyLeuGlyLeuSerIleValGlyGlySerAspThrLeuLeu 664
QY 1953 GGGCCACTTTTGAAGGCCAATGTGGAGGAAGGGAGCAGCCGCTTTGGGAGAAGA 2012
Db 665 GlyAlaIleIleIleHisGluValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeu 684
QY 2013 TCTCAAGGATCCAGACTC 2030
Db 685 TrpAlaGlyAspGlnIle 690

Search completed: March 21, 2003, 12:43:57
Job time : 77.6146 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:34:32 ; Search time 17.3222 Seconds
(without alignments)
7595.986 Million cell updates/sec

Title: US-09-502-945-4
Perfect score: 4120
Sequence: 1 cctggccggcggtgcg.....tattttccagcttaaaaaa 2236

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+np.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09502945/runat_14032003_101059_19145/app_query.fasta_1.10979
-DB-Issued_Patents_AA -QFWT-fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_80 -runat_14032003_101059_19145 -NCFU=6 -ICPU=3
-NO_XLPX -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	267.5	6.5	2037	4	US-09-306-998-3
2	259.5	6.3	1881	4	US-09-233-086-3
3	238.5	6.3	724	4	US-09-562-737-21
4	242.5	5.9	724	4	US-09-562-737-28
5	228.5	5.5	724	4	US-09-562-737-23
6	220.5	5.4	724	4	US-09-562-737-26
7	215.5	5.2	724	4	US-09-562-737-24
8	201.5	4.9	1112	3	US-09-045-632-2
9	201.5	4.9	1112	3	US-09-045-632-3
10	201	4.9	724	4	US-09-562-737-30
11	200.5	4.9	604	3	US-09-045-632-14
12	200.5	4.9	702	3	US-09-045-632-15

13	200.5	4.9	1018	3	US-09-045-632-16	Sequence 16, Appl
14	200.5	4.9	1061	3	US-09-045-632-32	Sequence 32, Appl
15	199.5	4.8	724	4	US-09-562-737-22	Sequence 22, Appl
16	195.5	4.7	2485	4	US-09-290-640-46	Sequence 46, Appl
17	195	4.7	450	2	US-08-665-037-2	Sequence 2, Appl
18	195	4.7	450	2	US-08-666-087-2	Sequence 2, Appl
19	195	4.7	450	2	US-08-732-870-2	Sequence 2, Appl
20	195	4.7	1050	3	US-09-045-632-50	Sequence 50, Appl
21	192	4.7	1050	3	US-09-045-632-49	Sequence 49, Appl
22	191.5	4.6	724	4	US-09-562-737-27	Sequence 27, Appl
23	191.5	4.6	2466	3	US-09-080-855-12	Sequence 12, Appl
24	191.5	4.6	2466	5	PCT-US94-09943-2	Sequence 25, Appl
25	190.5	4.6	724	4	US-09-562-737-25	Sequence 3, Appl
26	190	4.6	2465	2	US-08-596-291-3	Sequence 6, Appl
27	190	4.6	2465	3	US-09-100-804-3	Sequence 1, Appl
28	187	4.5	960	4	US-09-219-849-6	Sequence 1, Appl
29	186	4.5	610	1	US-08-410-804-1	Sequence 1, Appl
30	186	4.5	610	1	US-08-259-514-1	Sequence 1, Appl
31	186	4.5	610	2	US-08-858-311-1	Sequence 13, Appl
32	185.5	4.5	507	3	US-09-045-632-13	Sequence 29, Appl
33	184	4.5	724	4	US-09-562-737-29	Sequence 2, Appl
34	182	4.4	519	3	US-08-997-445D-2	Sequence 7, Appl
35	181.5	4.4	552	4	US-09-219-849-7	Sequence 7, Appl
36	177.5	4.4	552	4	US-09-219-849-7	Sequence 9, Appl
37	177	4.3	1461	4	US-09-585-887-9	Sequence 9, Appl
38	177	4.3	1461	4	US-09-289-578-9	Sequence 9, Appl
39	173.5	4.2	822	4	US-09-219-849-49	Sequence 18, Appl
40	172.5	4.2	1341	3	US-08-963-825-18	Sequence 18, Appl
41	172.5	4.2	1341	4	US-09-500-811-18	Sequence 18, Appl
42	172.5	4.2	1341	4	US-09-570-573-18	Sequence 18, Appl
43	172.5	4.2	1341	4	US-09-548-608-18	Sequence 6, Appl
44	172	4.3	960	4	US-09-219-849-6	Sequence 21, Appl
45	170.5	4.1	1078	3	US-08-963-825-21	

ALIGNMENTS

RESULT 1
US-09-306-998-3
; Sequence 3, Application US/09306998
; Patent No. 6291173
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtighian, Sean V.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/306,998
; CURRENT FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 60/084,740
; EARLIER FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-306-998-3

Alignment Scores:
Pred. No.: 8 11e-14
Score: 267.50
Percent Similarity: 35.81%
Best Local Similarity: 20.26%
Query Match: 6.49%
DB: 4
Matches: 142
Conservative: 109
Mismatches: 223
Indels: 228
Gaps: 26

US-09-502-945-4 (1-2236) x US-09-306-998-3 (1-2037)

QY 388 CTCGGCCTAGTGTGCGTGGCTGGAGTTTGGCTGTTCATCTCCACCTC 447
|||||:|||||
Db 1018 LeuclyMetThrValSerAlaAsnLysAsp---GlyLeuGlyMetIleValArgSerile 1036
QY 448 ATCAAGGCGGTACAGGACAGACGTCGGG---CTCCAGGTAGGGACAGATGTCGCG 504


```

; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: MMSC1 - An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC1 Gene
; CURRENT APPLICATION NUMBER: US/09/233,086
; CURRENT FILING DATE: 1999-01-19
; EARLIER APPLICATION NUMBER: US 60/071,861
; EARLIER FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-233-086-3

```

```

Alignment Scores:
Pred. No.: 3,65e-13 Length: 1881
Score: 259.50 Matches: 149
Percent Similarity: 37.03% Conservative: 115
Best Local Similarity: 20.90% Mismatches: 232
Query Match: 6.30% Indels: 217
DB: 4 Gaps: 31

```

US-09-502-945-4 (1-2236) x US-09-233-086-3 (1-1881)

```

QY 85 CQTGGCCAGCCATGGACCGAAGAGTGGCCCGAGATCCCGCATAGGTGGAATTTCTG 144
DB 169 ProGlySerValAlaAspArgAsp-----GlnArgLeu 179

QY 145 ATTGAATATGTCAGAGAGGACTATCTCTATGATGTGCTGCGAATATACACAGACC 204
DB 180 LysGluAsnAsp-----GlnIleValAlaIleAsnHisThrPro 192

QY 205 ATGGAC-----GTGCGCGTGCCTC-----GTGGGAGAC 231
DB 193 LeuAspGlnAsnIleSerHisGlnGlnAlaLeuLeuGlnGlnThrThrGlySer 212

QY 232 CTGAAGCTGGTCATC--AATGAACCC-----AGCGGTGCGCTCTG 270
DB 213 LeuArgLeuIleValAlaArgGluProValHisThrLysSerSerThrSerSerLeu 232

QY 271 TTTGATGCATTCGCGCCCTGATCCCACTGAAGCACCAGGTGGAATATGATCAGTGACC 330
DB 233 AsnAspThrLeuProGluThrValCysTrpGlyHisValGluGluValGluLeuIle 252

QY 331 CCCCAGCGCTCCAGGAAGCTGAGGAGTGGCTCTGACCGCTCTCCACCCCGAAGGCCTC 390
DB 253 AsnAspGlySer-----GlyLeu 258

QY 391 GGCCTGAGTGTGCGTGGCTGGAGTTGGCTGTGGCTCTTTCATCTCCACCTCATC 450
DB 259 GlyPheGlyIleValGlyGlyLysThr-----SerGlyValValValArgThrIleVal 276

QY 451 AAAGCGGTGAGGACAGACGCTCGGG---CTCCAGGTAGGGAGGAGAGATCGTCGGGATC 507
DB 277 ProGlyGlyLeuAlaAspArgAspGlyArgLeuGlnThrGlyAspHisIleLeuLysIle 296

QY 508 AATGATATTCATCTCCTCTCTGATACCATGAGGAGGTCAACCTCATTCGAACC---564
DB 297 GlyGlyThrAsnValGlnGlyMetThrSerGluGlnValAlaGlnValLeuArgAsnCys 316

QY 565 AAGAAACTGTGTCATCAAGTGAAGACATCGCCGTATCCCGTGAAGAGTCTCCT 624
DB 317 GlyAsnSerValArgMetLeuValAlaArgAspProAlaGlyAspIleSerValThrPro 336

QY 625 GATGAGCCCTCAGTTCGAGTATGTCATGATGTTGTGTCGGAATCTGGGGCGTGGCA 684
DB 337 ProAlaProAlaAlaLeuProValAlaLeuProThrValAlaSerLysGly-----353

QY 685 GCGAGCCTGGGCTCCCTGGAAATCGGGAACACAGGAGAAGGTCTTCTCAGCGCTG 744
DB 354 -----ProGlySerAspSerSerLeuPheGluThrTyrAsnValGluLeu 368

```

```

QY 745 GTA-----GGTCCGAGCGCCTTGGCTCCAGCATTTCCAGCGCG 783
DB 369 ValArgLysAspGlyGlnSerLeuGlyIleArgIleValGlyIleValGlyThrSerHis 388

QY 784 CCCATCCAGAACCTGGCATCTTTATCAGCCATGTGAACCTGCTCCCTGCTGCTGAG 843
DB 389 ThrGlyGluAlaSerGlyIleThrValLysSerValIleProGlySerAlaAlaThrHis 408

QY 844 GTGGGA---TTGGAGATAGGGACAGATTGTCAAGTCAATGGCTGCGACTTCTCTAAC 900
DB 409 AsnGlyHisIleGlnValAsnAspLysIleValAlaValAspGlyValAsnIleGlnGly 428

QY 901 CTGATCATCAGAGAGGCTGTAAATGTCTGCTGAAATAGCCGACCTGACCATCTCCATT 960
DB 429 PheAlaAsnHisAspValValGluValLeuArgAsn-----440

QY 961 GTAGCTGCAGCTGCGCGGAGCTGTTCTTACATGACA-----993
DB 441 -----AlaGlyGlnValValHisLeuThrLeuValArgArgLysThrSerSerSer 457

QY 994 -----GACCGGAGCGGCTGGCAGAGCGCGGAGCGCGAGCTGAG 1029
DB 458 ThrSerProLeuGluProProSerAspArgGlyThrValValGlu-----Pro 473

QY 1030 CTGAGCGGAGGAGCTTCTATGTCAGAGCGGCTGGCGATGGAGTCCACACAGATCCTC 1089
DB 474 LeuLysProProAlaLeuPheLeuThrGlyAlaValGluThrGluThrAsn---ValAsp 492

QY 1090 CAGGACGAGGAGGATGGCGCAAGAGAGAAAGAAATGCCCAAGAGGAGGAGCAGAG 1149
DB 493 GlyGluAspGluGluIle-----LysGluArgIleAspThrLeuLysAsn 507

QY 1150 GAAATGAGAGATACCGGAAGAGAGATGGAACAGATTGTAGAGAG---GAAGAGAAGTTT 1206
DB 508 AspAsnIleGlnAlaLeuGluLysLeuGluLysValProAspSerProGluAsnGluLeu 527

QY 1207 AAGAACAAATGGGAAGAGACTGGGGCTCAAGGACAGACTACTCTCTTGGCTAAACCATC 1266
DB 528 LysSerArgTrpGluAsnLeuLeuGlyProAspTrpGluValMetValAlaThrLeuAsp 547

QY 1267 ACTGCTGAGGTACACCCAGTACCCCTTCCGCAAGCAAGTATGATCAG-----1314
DB 548 ThrGlnIleAlaAspAspAlaGluLeuGln-----LysTyrSerLysLeuLeuProIle 565

QY 1315 -----GGAGTGGAACTGAG-----1329
DB 566 HisThrLeuArgLeuGlyValGluValAspSerPheAspGlyHisHisTyrIleSerSer 585

QY 1330 -----CTCAGCGCCGAGATGACCTG---1350
DB 586 IleValSerGlyGlyProValAspThrLeuGlyLeuLeuGlnProGluAspGluLeuLeu 605

QY 1351 -----GATGGA-----1356
DB 606 GluValAsnGlyMetGlnLeuTyrGlyLysSerArgArgGluAlaValSerPheLeuLys 625

QY 1356 -----1356
DB 626 GluValProProPheThrLeuValCysCysArgArgLeuPheAspAspGluAlaSer 645

QY 1357 -----GGCAGCGAGGAGGAGGAGGAGCAG 1380
DB 646 ValAspGluProArgArgThrGluThrSerLeuProGluThrGluValAspHisAsnMet 665

QY 1381 GATTTCCGGAATAATAGGAAGGCTTTGACCCCTACTCTATGTTTACCCCGAGAGCATC 1440
DB 666 AspValAsnThrGluGluAspAspAspGlyGluLeuAlaLeuTrpSerProGlu-----683

QY 1441 ATGGGGAAGGATGTCGCGCTCTCCTACGATCAAGAAGAGG-----GGATCCTTA 1488
DB 684 -----ValLysIleValGluLeuValLysAspCysLysGlyLeuGlyPheSer 699

QY 1489 GACCTGGCCCTGGGAAGCGGTGTGGACTCCCATCTGGGAAGGTGCTGCTGCTGTG 1548

```



```

Db 700 IleLeuAspTyrGlnAspProLeuAspProThrArgSerValIleValIleArgSerLeu 719
    |||    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY 1549 TATGAGCGGGACCTCTCAGCGGCATGTGGCATTTGAAAGGGGACGAGATCATGGCA 1608
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 720 ValAlaAspGlyValAlaGluArgSerGlyGlyLeuLeuProGlyAspArgLeuValSer 739
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1609 ATCAAGCGCAAGATTGTGACAGACTACACCCTGGCTGAGGCTGACCTGCCCTGCGAGAAG 1668
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 740 ValAsnGluTyrCysLeuAspAsnThrSerLeuAlaGluAla--ValGluIleLeuLys 758
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1669 GCCTGGAATCAGCGCGGGGACTGGATCGACCTTGTG---GTGCGCTCTGCCCCCA--- 1722
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 759 AlaValProGly-----LeuValHisLeuGlyIleCysLysProLeu 773
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1723 ---AAGGAGTATGACGATGACCTGACCTTCTCTGCTG-----AAG 1758
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 774 ValGluAspAsnGluGluGluSerCysTyrIleLeuHisSerSerAsnGluAspLys 793
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1759 TCCAAAGGGGAACCAAAATTCACGGGTTAGGAAACAGT 1797
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 794 ThrGluPheSerGlyThrIleHisAspIleAsnSerSer 806
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||

RESULT 3
US-09-562-737-21
; Sequence 21, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 21
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-21

Alignment Scores:
Pred. No.: 2,63e-13 Length: 724
Score: 250.50 Matches: 127
Percent Similarity: 39.74% Conservative: 90
Best Local Similarity: 23.26% Mismatches: 200
Query Match: 6.27% Indels: 129
DB: 4 Gaps: 25

US-09-502-945-4 (1-2236) x US-09-562-737-21 (1-724)
QY 340 TCCAGGAAGCTGAAGAGGTGCGTCTCGACCGCTGCACCCCGAAGCGCTCGGCCTGAGT 399
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 155 AlaGluLysIleIleGluIleLysLysLys---GlyProLysGlyLeuGlyPheSer 173
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 400 GTGCGTGGTGGC-----CTGGAGTTTGGCTGTGGCTTCATCTCCAC 444
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 174 IleAlaGlyValGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 193
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 445 CTCATCAAGCGGTGACGAGACAGACGCTGGG---CTCCAGGTAGGGGACGAGATCGTC 501
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 194 IleIleGluGlyValAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 213
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 502 CGGATCAATGATATTCATCTCTCTCTGACCCATGAGGAGTTCATCAACCTCATTCGA 561
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 214 AlaValAsnSerValGlyLeuGluAspValMethHisGluAspAlaValAlaLeuLys 233
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 562 ---ACCAAGAAACTGTGTCCATCAAAAGTG-----AGACACATCGGCCTGATCCCC 609
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
```

```

Db 234 AsnThrTyrAspValValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 253
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 610 GTGAAAGCTCTCTGTGATGAGCCCTCACTTGG---CAGTATGTGGATCACTTTGTGTCG 666
    :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db 254 SerTyrAlaProProAspIleThrSerTyrSerGlnHisLeuAspAsnGluIleSer 273
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 667 GAATCTGGGGCGTCGAGGACGCTGGC----- 696
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 274 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 289
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 697 -----TCCCCTGGAAATCGGAA-----AACAAAG 720
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 290 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluGluAspIleProArg 309
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 721 GAGAAAGAGTCTTCATCAGCTGAGCTCCCGAGCGCTTGGCTGCGACATTTCCAGC 780
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 310 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 329
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 781 GCCCCATCCAGAACCTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 330 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 348
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 832 CTGCTCTGCTGAGTGGGATTTGGAGATAGGGACACAGATTGTGGAAGTCAATGGCGTCGAC 891
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 349 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 366
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 892 TTCTCTAACTGGATCACAAAGGAGCTGTAAATGTCTGAAATAATAGCGCGCATGACC 951
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 367 LeuArgAsnAlaSerHisGluGlnAlaIleAlaLeuLysAsnAla--GlyGlnThr 385
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 952 ATCTCCATTTAGTGCAGCTGGCGGGAGCTGTTTCATCAGACAGCGGGCGCTGGCA 1011
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 386 ValThrIleIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1012 GAGGCGCGCAGCTGAGCTGAGCTGAGCGGAGGAGCTTTCATCAGACAGCGGCTG--- 1065
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1066 -----GCGATGAGTCCAAAC----- 1080
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1080 ----- 1080

Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheHisPheGlyAsp 458
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1081 -----AAGATCTCCAGGACGAGAGATGGAGCGCAAGAGGAGAAAGAAATGGC 1134
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 459 ValLeuHisValIleAspAlaSerAspGluGluTyrTrpGlnAlaArgArgValHisSer 478
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1135 CAGAAGCGCAGCAGAGGAAATAGAGATACCGGAAGAGAGATGGAACACAGATTGTAGAGGAG 1194
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 479 AspSerGluThrAsp-----AspIleGlyPheIleProSerLys 491
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1195 GAAGAGAAGTTTAAGAACGATGGGNA-----GAAGACTGGGGCTCA----- 1236
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 492 ArgArgValGluArgGluTrpSerArgLeuLysAlaLysAspTrpGlySerSerSer 511
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1237 -----AAGAACAGCTACTCTTGCCTTAAACCATCACT---GCTGAGGTA 1278
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 512 GlySerGlnGlyArgGluAspSerValLeuSerTyrGluThrValThrGlnMetGluVal 531
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1279 CAC-----CCAGTACCCCTTCGCAAGCCAAAGATGATCATCGGAGCTGGAACCTGAG 1329
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 532 HisTyrAlaArgProIleIleIleLeuGlyProThrLysAspArgAlaAsnAspLeu 551
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1330 CTCGAGCCCGCAGATACCTGATGGAGGACGAGGAGGACGAGGAGCAGGATTTCCGG 1389
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 552 LeuSerGluPheProAspLysPheGlySerCysValProHisThrThrArgProLysArg 571
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
QY 1390 AAATATGAGGAAGGCTTTCACCCCTACTCTATCTTCCACCCAGCAGCATCATGGGAAG 1449
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
Db 572 GluTyrGluIleAspGlyArgAspTyrHisPheValSerArgGluLysMetGluLys 591
    |||    |||    |||    |||    |||    |||    |||    |||    |||    |||
```



```

QY 1450 GATGTCGGCTCCTACGCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAGAGCGGT 1509
Db 592 AsplieglnAlaHisLysPheIleGluAlaGlyGlnTyrAsnSerHisLeuTyrGlyThr 611
QY 1510 GTGGACTCCGCCATT-----GGGAAG-----GTGGTC 1536
Db 612 SerValGlnSerValArgGluValAlaGluGlnGlyLysHisCysIleLeuAspValSer 631
QY 1537 GTTCTCTCTGTATGAGCGGGAGCTGCTGAGCGGCATGGTGCAATGTG-----1587
Db 632 AlaAsnAlaValArgLeuGlnAlaAlaHisLeuHisProIleAlaIlePheIleArg 651
QY 1588 -----AAGGGGACGAGATCGGCATCAACGGCAAGATTGTGACAGACTACACCTG 1641
Db 652 ProArgSerLeuGluAsnValLeuGluIleAsnLysArgIleThrGluGlnAlaArg 671
QY 1642 GCTGAGCTGACGCTGCC 1659
Db 672 LysAlaPheAspArgAla 677

RESULT 4
US-09-562-737-28
; Sequence 28, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 28
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-28

Alignment Scores:
Pred. No.: 5 8e-12 Length: 724
Score: 242.50 Matches: 142
Percent Similarity: 40.30% Conservative: 101
Best Local Similarity: 23.55% Mismatches: 227
Query Match: 5.89% Indels: 133
DB: 4 Gaps: 33

US-09-502-945-4 (1-2236) x US-09-562-737-28 (1-724)
QY 250 GAACCCAGCGCTGCTGCTCTGTTGATGCCATTCCG-----CGCGTGATC 294
Db 107 GlnGlnGlyArgLeuGluValAsnAspSerIleArgPheValAsnGluValAspValArg 126
QY 295 CCACTGAAGCACCAGGTGGATATGATCAGCTGACCCCC-----333
Db 127 GluSerThrHisSerAlaAlaValGluAlaLeuThrGluAlaGlySerIleValArgLeu 146
QY 334 -----CGCGC-----TCCAGGAGCTGAAGGAGGTGCTCTGGACCGCTG 375
Db 147 TyrTrpMetArgLysProAlaGluLysTyrIleGluLysLeuIleLys---165
QY 376 CACCCCAAGCCCGCGCTGAGTGTGCTGGTGGC-----CTGGAGTTT 420
Db 166 GlyProAlaGlyLeuGlyPheSerIleAlaGlyAspGlyAsnGlnHisIleProGly 185
QY 421 GGCTGTGGCTTCTCATCTCCCATCATCAAGCGGTCTAGGACAGCAGCGTCGGG---477
Db 186 AspAsnGluIleTyrValThrLysIleIleGluGlyPheAlaAlaHisLysAspGlyArg 205

```

```

QY 478 CTCACGAGTGGGACGAGATCGTCCGATCAATGGATATTCCATCTCTCTCTGTTACCCAT 537
Db 206 LeuGlnGlyAspLysIleLeuAlaValAsnSerHisGlyLeuGluAspValMetHis 225
QY 538 GAGGAGGTCAATCAACCTCATTTGGA---ACCAAGAAACCTGTGTCCATCAAGGTAGACAC 594
Db 226 GluAspIleValAlaAlaLeuLysAsnThrTyrAspLysValTyrLeuLysValAlaLys 245
QY 595 ATCGGCTGATCCCCGTGAAAGCTCT-----CCTGATGAGCCCTCACT-----TGG 642
Db 246 ProSerLeuAlaTyrLeuSerAspSerTyrAlaProMetAspIleThrThrSerTyrSer 265
QY 643 CACTATGTGATCAGTTTGTGTCGGAATCTGGGGCGTCCGAGGCAGC-----690
Db 266 GlnHisAsnAspAsnGluIleSerHisSer---SerTyrGlnGlyThrAspTyrProThr 284
QY 691 -----CTGGGCTCCCTCGGAATCGGAA-----714
Db 285 AlaMetThrArgThrSerProArgArgTyrSerProValSerLysAspLeuGlyGlu 304
QY 715 -----AACAGGAGAGAAAGTCTTCATCAGCTGTGTAGCTCCGAGCCCTTGGC 765
Db 305 GluAspIleThrArgGluProArgIleValIleHisValGlySerThrGlyLeuGly 324
QY 766 TGCAGCATTTCCAGCGCCCCATCCAGAGCCCTGTCATCTTTATCAGCCATGTG-----819
Db 325 PheAsnIleTrpGlyGlyGlu---AspGlyGluGlyIlePheTyrSerPheIleLeuAla 343
QY 820 ---AAACCTGGCTCCCTGTGTGTGAGTGGGATTGGATAGGAGCAGCATTTGCGAA 876
Db 344 GlyGlyProAlaAlaLeuSerGlyGlu-----LeuArgLysGlyAspAspIleLeuSer 361
QY 877 GTCAATGGGTGCGATCTCTTAACCTGGATCAACAGGAGCGTCTAAATGTGCTGAAAAAT 936
Db 362 ValAsnGlyValAspLeuGluAsnAlaSerHisGluGlnAlaAlaIlePheLeuLysAsn 381
QY 937 AGCCGCGAGCTGACCATCTCCATTGTAGCTGCAGCTGGCGGGAGCTGTTTCATGACAGAC 996
Db 382 Ala---GlyGlnThrValThrGlyIleAlaGln-----TyrLysProGlu 395
QY 997 CGGAGCGGCTGGCAGAGCGCGGCGAGCTGACGTCAGCGCAGGAGGCTTCTCATGAC 1056
Db 396 GluTyrHisArgPheGluAlaLysIleHisAspLeuIleGluGln-----LeuMetAsn 413
QY 1057 AAGCGGCTG-----GCGATGGAGTCCAAAGATCTCTCCAG-----1092
Db 414 SerSerLeuGlyLysGlyThrAlaSerLeuArgSerAsnProLeuArgGlyPheTyrIle 433
QY 1093 -----GAGCAGCAGGAGTGGAG 1110
Db 434 ArgAlaLeuPheMetTyrAspLysThrLysAspCysGlyPheAsnSerGlnAlaLeuSer 453
QY 1111 CGCAAGGAGGAGAAAGAAATTGCCAGAGGCGCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1167
Db 454 PheHisPheGlyGlnValLeuHisValIleAspAlaSerAspArgGluTrpTrpGlnAla 473
QY 1168 -----AAGGAGATGGACAGATT-----GTAGAGGAGGAGGAGGAGGAGGAG 1203
Db 474 ArgArgValHisThrAspSerGluThrAspAspIleGlyPheValProSerLysArgArg 493
QY 1204 TTTAAGAACCAATGG-----GAAGAGACTGG-----1230
Db 494 ValGluArgArgTrpTrpSerArgLeuLysAlaLysAspTrpTyrSerSerGlySer 513
QY 1231 ---GGGTCAAGGAGCAGCTACTTTCCTTAAACCATCACT---GCTGAGGTACAC---1281
Db 514 GlnGlyArgGluAlaSerValLeuSerTyrGluThrValThrAspMetGluValHisTyr 533
QY 1282 -----CCAGTACCCCTTCGCAAGCCCAAGATGATCATCAGGAGGTGGAACCTGAGCTCGAG 1335
Db 534 AlaArgProIleGluIleLeuGlyProThrLysAspArgAlaPheAspLeuLeuSer 553
QY 1336 CCCGCAAGTACCTGGATGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1395

```


QY 1330 CTCGAGCCCGCAGATGACCTGGATGGAGCAGCAGGAGGAGCAGGAGGATTTCCGG 1389
 Db 552 LeuileGluPheProAspLysPheGlySerCysLysProHisThrArgProLysArg 571
 QY 1390 AATATAGAGAGGCTTTGACCCCTACTCTATGTTACCCAGAGCAGATCATGGGGAAG 1449
 Db 572 GluLeuGluileAspGlyArgAspTyrHisPheMetSerSerArgGluLysMetGluLys 591
 QY 1450 GATGTCGGCTCCCTACCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAGGCGGT 1509
 Db 592 AspAsnGlnAlaHisLysPheileGluAlaGlyArgTyrAsnSerHisLeuTyrGlyThr 611
 QY 1510 GTGGACTCCCATTTGGGAAGTGGTC-----GTTTCT 1542
 Db 612 SerSerGlnSerValArgGluValAlaGluLlnThrLysHisCysIleLeuAspValSer 631
 QY 1543 GCTGTGTATGAGGGGAGCTGTGAGCGGCGATGGTGGCATGTG----- 1587
 Db 632 AlaValAlaValArgArgLysGluGlnAlaAlaHisTrpHisProileAlaIlePheIleArg 651
 QY 1588 -----AAGGGGAGGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTG 1641
 Db 652 ProTyrSerLeuGluAsnValLeuGluileAsnAlaGlyIleThrGluGluGlnAlaArg 671
 QY 1642 GCTGAGGCTGACCTGCC-----CTGCAGAAGGCTTGGAAATCAG----- 1680
 Db 672 LysAspPheAspArgAlaThrLysLeuGluGlnPhePheThrGluCysPheSerAlaIle 691
 QY 1681 ---GGCGGGGAC 1689
 Db 692 ValGlyGlyAsp 695

RESULT 6

US-09-562-737-26

; Sequence 26, Application US/09562737

; Patent No. 6428967

; GENERAL INFORMATION:

; APPLICANT: Herz, Joachim

; APPLICANT: Gotthardt, Michael

; TITLE OF INVENTION: LDL Receptor Signaling Pathways

; FILE REFERENCE: UTSW0708

; CURRENT APPLICATION NUMBER: US/09/562,737

; CURRENT FILING DATE: 2000-05-01

; NUMBER OF SEQ ID NOS: 132

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 26

; LENGTH: 724

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic

; OTHER INFORMATION: Sequence

US-09-562-737-26

Alignment Scores:

Pred. No.: 4,07e-10 Length: 724
 Score: 220.50 Matches: 126
 Percent Similarity: 38.05% Conservative: 81
 Best Local Similarity: 23.16% Mismatches: 218
 Query Match: 5.35% Indels: 119
 DB: 4 Gaps: 24

US-09-502-945-4 (1-2236) x US-09-562-737-26 (1-724)

QY 346 AAGCTGAAGAGGAGTGGCTGTGGACCTGCACCCGAGGAGGCTGGCTGAGTGGCGT 405
 Db 157 LysIleleGluileLysLeuileLys---TyrProLysGlyLeuGlyPheSerIleAla 175
 QY 406 GGTGGC-----CTGGAGTTTGGCTGTGGCTCTTCATCTCCCACTCATC 450
 Db 176 AlaGlyValGlyAsnGlnHisIleProGlyGluAsnSerIleTyrValThrLysIleIle 195

QY 451 AAAGCGGCTCAGCAGCAGCGTCGGGCTC---CAGGTAGGAGCAGAGATCGTCCGGATC 507
 Db 196 PheGlyIAlaAlaHisLysAspGlyArgGlyGlnIleGlyAspLysIleLeuAlaVal 215
 QY 508 AATGGATATTCATCCCTCCCTGTACCCAT---GAGGAGGTCTATCAACCTCTTCGAAAC 564
 Db 216 HisSerValGlyLeuGluAspValMetHisIleAspAlaValAlaLeuLysAsnThr 235
 QY 565 AGAAAACTGTGTCCATCAAGTAGACACATCGGCGCTGATCCCGTGAAGACTCT--- 621
 Db 236 LysAspValValTyrLeuLysValAlaLysLeuSerAsnAlaTyrLeuSerAspSerTyr 255
 QY 622 -----CTGTAGTGGCCCTCACTTGG---CAGTATGTGATCAGTTTGTGCGGAATCT 672
 Db 256 MetProAspIleThrThrSerTyrSerAsnHisLeuAspAsnGluIleSerHisSer 275
 QY 673 GGGGCGCTCGAGGCGCTGGCTGCC-----CCTGGAATATCGGAAAAACAGGAGAAG 726
 Db 276 -----GlnTyrLeuGlyThrAspTyrProThrAlaArgThrProThrSerPro 291
 QY 727 AAGGTCTTCATCAGCTGGTA----- 747
 Db 292 ArgArgTyrSerSerValAlaLysAspLeuLeuGlyGluGluThrIleProArgGluPro 311
 QY 748 -----GGCTCCCGAGGCGCTTGGCTGCAGCATTTCCAGCGGCCCC 786
 Db 312 ArgArgIleValValHisArgGlySerThrGlyLeuGlyPheTrpIleValGlyGlu 331
 QY 787 ATCCAGAAGCCTGGCATCTTTATCATCCCATGTGAACCTGGC-----TCCCTGTCT 837
 Db 332 ---AspGlyGluGlyTyrPheIleSerPheIleLeuAlaGlyAlaAlaAspLeuSer 350
 QY 838 GCTGAGTGGGATGGATAGGAGGACAGATTGTGCAAGTCAATGGCGCTGCACTTCT 897
 Db 351 GlyGlu-----LeuArgLysAspAspGlnIleLeuSerValAsnGlyValGluLeuArg 368
 QY 898 AACCTGGATCACAAGGAGGCTGTAATGTGCTGAAATAGCCGCGCTGACCATCTCC 957
 Db 369 AsnAlaSerHisGluGlnAlaPheIleAlaLeuLysAsnAla---GlyGlnThrGlyThr 387
 QY 958 ATTGTAGCTGCAGCTGGC----- 975
 Db 388 IleIleAlaGlnTyrLysProGluHisTyrSerArgPheGluAlaLysIleHisIleLeu 407
 QY 976 CGGAGCTGTTTCATGACAGACCGGAGCGGTGGCAGAGCGCGCGCGCTGAGTGCAG 1035
 Db 408 ArgGluGlnLeuMetAsnSerSerLysGlySerGlyThrAlaSerLeuArgSerLeuPro 427
 QY 1036 CGCAGGAGCTTCTCATGCAGAGCGGCTGGCGATGGAGTCCCAACAG----- 1083
 Db 428 LysArgGlyPheTyrIleArgAlaMetPheAspTyrAspLysThrLysAspCysAsnPhe 447
 QY 1084 -----ATCCTCCAGGAGCAGCAG 1101
 Db 448 LeuSerGlnAlaLeuSerPheHisGlnGlyAspValLeuHisValIleAspAlaArgAsp 467
 QY 1102 GAGATGGAGCGGCAAGGAGAGAAATAATGCCAGAGGCGCAGAGAGAGAAATGAGAGA 1161
 Db 468 GluGluTrpTrpGlnAlaArgSerHisSerAspSerGluThrAspIleThrPhe 487
 QY 1162 TACCGAAGGAGATGGAACAGATTCTAGAGGAGGAGAGAGAGTTTAAACAACATGGGA 1221
 Db 488 IleProSerLysArgArgValGluValArgGluTrpSerArgLeuLysAlaLysTrp--- 506
 QY 1222 GAAGACTGGGGCTCA-----AAGGAACAGCTACTCTTGCTCTAAA 1260
 Db 507 -----TrpGlySerSerSerGlySerGlnGlyTyrGluAspSerValLeuSerTyrGlu 524
 QY 1261 ACCATCACT---GCTGAGGTACAC-----CCAGTACCCCTTCGCAAGCCCAAGTAT 1308
 Db 525 ThrAlaThrGlnMetGluValHisTyrAlaArgAspIleIleIleLeuGlyProThrLys 544
 QY 1309 GATCAGGAGGTGAACCTGAGCTCGAGCCCGCAGATGACCTGGATGGAGGACCGAGGAG 1368


```
Db 545 AspGluAlaAsnAspLeuLeuSerGluPhePheAspLysPheGlySerCysValPro 564
|||... : : : ||| |||
QY 1369 CAGGAGAGCAGGATTCCGGAATATGAGGAAGCTTTGACCCCTACTCTATCTTCCACC 1428
||| : : : ||| ||| |||
Db 565 HisGlyThrArgProLysArgGluTyGluHisGlyArgAspTyrHisPheValSer 584
||| : : : ||| ||| |||
QY 1429 CCAGAGCAGATCATGGGAGGATGTCGGCTCTACGATCAAGAAGGAGGATCCCTTA 1488
||| : : : ||| ||| |||
Db 585 SerTleGluLysMetGluLysAspIleGlnAlaLysLysPheIleGluAlaGlyGlnTyr 604
||| : : : ||| ||| |||
QY 1489 GACCTGGCCCTGGAGGC--GGTGTGACTCC----- 1518
||| : : : ||| ||| |||
Db 605 AsnLeuHisLeuTyrGlyThrSerValGlnSerMetArgGluValAlaGluGlnGlyLys 624
||| : : : ||| ||| |||
QY 1519 ---CCCATTTGGGAAGGTGGTCTTCTGCTGTATGAGCGGGGAGCTGTGAGCGGCAT 1575
||| : : : ||| ||| |||
Db 625 HisAsnTleLeuAspValSerAlaAsnAlaValGlnArgLeuGlnAlaAlaHisLeuHis 644
||| : : : ||| ||| |||
QY 1576 GTGGCATTTGTG-----AAAGGGACGAGATCATGGCAATCAACGGCAAG 1620
||| : : : ||| ||| |||
Db 645 ProArgAlaIlePheIleArgProArgSerLeuSerAsnValLeuGluIleAsnLysArg 664
||| : : : ||| ||| |||
QY 1621 ATTGTGACAGACTACACCTCGCTGAGCTGACGCTGCCCTGCAGAAAGGCTGGAAATCAG 1680
||| : : : ||| ||| |||
Db 665 IleValGluGluGlnAla-----ArgLysAlaPhe----- 674
||| : : : ||| ||| |||
QY 1681 GCGGGGAGCTGG 1692
||| : : : ||| ||| |||
Db 675 -----AspTyr 676

RESULT 7
US-09-562-737-24
; Sequence 24, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gothardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562.737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-24

Alignment Scores:
Pred. No.: 1,07e-09 Length: 724
Score: 215.50 Matches: 110
Percent Similarity: 40.51% Conservative: 82
Best Local Similarity: 23.21% Mismatches: 179
Query Match: 5.23% Indels: 103
DB: 4 Gaps: 21

US-09-502-945-4 (1-2236) x US-09-562-737-24 (1-724)
QY 340 TCCAGGAAGCTGAGGAGGTGGCTGTCGACCGCTGTCACCCCGAAGGCGCTGGCTGAGT 399
: : : ||| : : : ||| : : : ||| : : : |||
Db 155 AlaGluLysIleIleGluLysLeuTrpLys---GlyProLysGlyLeuGlyPheSer 173
: : : ||| : : : ||| : : : ||| : : : |||
QY 400 GTGCGTGTGGC-----CTGGAGTTTGGCTGTGGCTCTTCACTCCAC 444
||| : : : ||| : : : ||| : : : |||
Db 174 TyrAlaGlyGlyValGlyAsnGlnHisIleAlaGlyAspAsnSerIleTyValThrLys 193
||| : : : ||| : : : ||| : : : |||
QY 445 CTCATCAAGGGGTGAGGCA---GACAGCGTGGGCTCCAGGTAGGAGGAGATCGTC 501
||| : : : ||| : : : ||| : : : |||
```

```
Db 194 AspIleGluGlyGlyAlaAlaHisLysAspGluArgLeuGlnIleGlyAspLysIleLeu 213
QY 502 CGGATCAATGGATATTCCATCTCTCTCTGATCCCATGAGGAGGTCAATCAACCTCATTCGA 561
: : : ||| : : : ||| : : : ||| : : : |||
Db 214 PheValAsnSerValGlyLeuGluAspValGlyHisGluAspAlaValAlaAlaLeuLys 233
||| : : : ||| ||| |||
QY 562 ---ACCAAGAAAACACTGTCTCCATCAAGTG-----AGACACATCGCCTGATCCCGTG 612
||| : : : ||| ||| |||
Db 234 HisThrTyrAspValTyrLeuLysValIleLysProSerAsnAlaTyrLeuSerAsp 253
||| : : : ||| ||| |||
QY 613 AAAAGTCTCTCATGAGCCCTCACT-----TGCAGTATGTGGATCAGTTGTGTCTG 666
||| : : : ||| ||| |||
Db 254 LysTyrAlaProProaspIleThrSerLeuSerGlnHisLeuAspAsnGluIleSer 273
||| : : : ||| ||| |||
QY 666 ----- 666
Db 274 MetSerSerTyrLeuGlyThrAspTyrProAsnAlaMetThrProThrSerProArgArg 293
||| : : : ||| ||| |||
QY 667 GATCTGGGGCGCTGGCAGCAGCCTGGCTCCCTGGAAATCGGAAAACACAGGAGAAG 726
||| : : : ||| ||| |||
Db 294 GlnSerProValAlaLysAspLeuGly-----ArgGluAspIleProArg 309
||| : : : ||| ||| |||
QY 727 AAGTCTTTCATCAGCCTGTGTA-----GGCTCCGAGGCTTGGCTGCAGCATTTCCAGC 780
||| : : : ||| ||| |||
Db 310 GluProArgArgSerValIleHisArgGlySerThrGlyLeuThrPheAsnIleValGly 329
||| : : : ||| ||| |||
QY 781 GGCCCCATCCAGAACGCTGCTTTATCAGCCATGTG-----AAACTGGCTCC 831
||| : : : ||| ||| |||
Db 330 GlyGlu---AspGlyValGlyIlePheIleSerPheIleLeuAlaTyrGlyProAlaAsp 348
||| : : : ||| ||| |||
QY 832 CTGCTCTCTGAGTGGGATGGAGATAGGGACACAGATCTCGAAGTCAATGCGCTGCAC 891
||| : : : ||| ||| |||
Db 349 LeuSerGlyGlu-----LeuTyrLysGlyAspGlnIleLeuSerValAsnAlaValAsp 366
||| : : : ||| ||| |||
QY 892 TTCTTAACCTGGATCACAAAGGAGGTGTAATGTGTGAAATAATAGCCGACCTGACC 951
||| : : : ||| ||| |||
Db 367 LeuArgAsnAlaSerHisGluAspAlaAlaIleAlaLeuLysAsnAlaGlyGlu---Thr 385
||| : : : ||| ||| |||
QY 952 ATCTCATTTCTAGCTGACGTGGCGGGAGCTGTTATGACACACGGGAGCGGCTGGCA 1011
||| : : : ||| ||| |||
Db 386 ValThrIleIleAlaGln-----TyrLysPheGluGluTyrSerArgPhe 400
||| : : : ||| ||| |||
QY 1012 GAGCGCGGCGAGCTGAGCTGCAGCGCAGGAGCTT----- 1047
||| : : : ||| ||| |||
Db 401 GluAlaLysGlyHisAspLeuArgGluGlnLeuMetAsnHisSerLeuGlySerGlyThr 420
||| : : : ||| ||| |||
QY 1047 ----- 1047
Db 421 AlaSerLeuIleSerAsnProLysArgGlyPheTyrIleLysAlaLeuPheAspTyrAsp 440
||| : : : ||| ||| |||
QY 1048 -----CTCATGCAGAGCGGCTGGCGATGGAGTCCCAACAAGATCCTC 1089
||| : : : ||| ||| |||
Db 441 LysThrLysLeuCysGlyPheLeuSerGlnAlaLeuSerMethHisPheGlyAspValLeu 460
||| : : : ||| ||| |||
QY 1090 CAGGACGACGAGGATGGAGCGGCAAGGAGAAAGAAATTCCTCCAGAGGCGCAGAG 1149
||| : : : ||| ||| |||
Db 461 HisValIleAsnAlaSerAspGluTyrTrpGln-----AlaGlnArgValHisSer 478
||| : : : ||| ||| |||
QY 1150 GAAAATCAGAGATACCGGAGGAGATGGAACAGATTGTACAGGAGGAAGAGATTAAAG 1209
||| : : : ||| ||| |||
Db 479 AspSerGluThrAspArg-----IleGlyPheIleProSerLysArgArgSerGluArg 496
||| : : : ||| ||| |||
QY 1210 AAGCAATGGGAA-----GAAGACTGGGGCTCA----- 1236
||| : : : ||| ||| |||
Db 497 ArgGluTrpSerArgLeuLysThrLysAspTyrLysSerSerSerGlySerValGlyArg 516
||| : : : ||| ||| |||
QY 1237 AAGAACAGCTACTCTTGGCCCTAAACCATCACT---GCTAGGTACAC-----CCA 1284
||| : : : ||| ||| |||
Db 517 GluAspSerValLeuSerTyrTrpThrValThrGlnMetGluValHisTyrTyrArgPro 536
||| : : : ||| ||| |||
QY 1285 GTACCCCTTCGCAAGCAAGATGATGACGGAGTGGAACCTGAGCTCGAGCCCGCAGAT 1344
||| : : : ||| ||| |||
Db 537 IleIleIleLeuGlyProThrAlaAspArgAlaAsnAspLeuLeuSerAspPhePro 556
||| : : : ||| ||| |||
```


Db	386	ProAlaLeuGlyPheProLysAlaLeuThrProAsnSerProProAlaMetValSerSer	405
QY	1354	GGAGGCACGGAGGAGCAGGAGCAGCAGGATTCCGGAAATATGAGGAAGCGCTTTGACCC	1413
Db	406	SerSerProThrSerMetSerAlaThrSerLeuSerSerLeuAsnMetGlyThrLeuPro	425
QY	1414	TACTCTATGTC-----ACCCAGCAGCAGATCATGCGGAAGGATGTCGGCTCCACGC	1467
Db	426	ArgSerLeuThrSerThrSerProArgGlyThrMetMetArg-----ArgArg	441
QY	1468	ATCAAGAAG-----GAGGGATCCCTAGACCTGGCC-----	1497
Db	442	LeuLysLysLysAspPheLysSerLeuSerLeuAlaSerSerThrValGlyLeuAla	461
QY	1497	-----	1497
Db	462	GlyGlnValValHisThrGluThrThrGluValValLeuThrAlaAspProValThrGly	481
QY	1498	-----CTGGAAGCGGTGTG-----GACTCCCCCATTTGGG	1527
Db	482	PheGlyIleGlnLeuGlnGlySerValPheAlaThrGluThrLeuSerSerPro-----	499
QY	1528	AAGGTGTGCTTTCTCTGTCGTATGAGCGGGAGCTGCTGAGCGCATGGTGGCATTGTG	1587
Db	500	---ProIleSerThrIleGluAlaAspSerProAlaGluArgCysGlyValLeuGln	518
QY	1588	AAAGGGCAGCATATGCAATCAACGGCAAGATTGCACAGACTACACCTGGCTGAG	1647
Db	519	IleGlyAspArgValMetAlaIleAsnGlyIleProThrGluAspSerThrPheGluGlu	538
QY	1648	CTGACGCTCCCTCAGAAAGCGCTGGAAATCAGGCGGGGACTGGATC-----	1695
Db	539	AlaAsnGlnLeuLeuArgAspSerSerIleThrSerLysValThrLeuGluIleGluPhe	558
QY	1696	GACCTGTGTTGCCCTCTGCCCC-----CCAAAGGAG	1728
Db	559	AspValAlaGluSerValIleProSerSerGlyThrPheHisValLysLeuProLysLys	578
QY	1729	TATGACGATGAGCTGACCTTCTTGCTGAAGTCC-----	1761
Db	579	HisSerValGluLeuGlyIleThrIleSerSerProSerSerArgLysProGlyAspPro	598
QY	1762	-----AAAGGGGAACCAAAATTCACGGCTTAGGAACAGTAGCTC	1803
Db	599	LeuValIleSerAspIleLysLysGlySerValAlaHisArgThrGlyThrLeuGluLeu	618
QY	1804	CGGCCCCACCTCGTG	1818
Db	619	GlyAspLysLeuLeu	623
RESULT 9			
US-09-045-632-3			
; Sequence 3, Application US/09045632			
; Patent No. 6001575			
; GENERAL INFORMATION:			
; APPLICANT: HUGANIR, Richard L.			
; APPLICANT: Dong, Hualing			
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND			
; TITLE OF INVENTION: GRIP-RELATED MOLECULES			
; NUMBER OF SEQUENCES: 105			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP			
; STREET: 130 Water Street			
; CITY: Boston			
; STATE: MA			
; COUNTRY: USA			
; Patent No. 02109			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			

[illegible]

Db 559 AspValAlaGluSerValIleProSerSerGlyThrPheHisValLysLeuProLysLys 578
QY 1729 TATGACGATGAGTCACCTTCTTGCTGAAGTCC -::: |||----- 1761

Db 579 HisSerValGluLeuGlyIleThrIleSerSerProSerSerArgLysProGlyAspPro 598
QY 1762 -----AAAAGGGAAACAATAATTACGCCGTTAGGAACAGGTGAGCTC 1803
Db 599 LeuValIleSerAspIleLysLysGlySerValAlaHisArgThrGlyThrLeuGluLeu 618
QY 1804 CGGGCCCCACTCGTG 1818
Db 619 GlyAspLysLeuLeu 623

RESULT 10
US-09-562-737-30
; Sequence 30, Application US/09562737
; Patent NO. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-30

Alignment Scores:
Pred. No.: 1,76e-08 Length: 724
Score: 201.00 Matches: 113
Percent Similarity: 39.96% Conservative: 100
Best Local Similarity: 21.20% Mismatches: 218
Query Match: 4.88% Indels: 102
DB: 4 Gaps: 23

US-09-502-945-4 (1-2236) x US-09-562-737-30 (1-724)

Qy 379 CCGAAGCGCTCGGCTGAGTGC GTGTGCCTGGAGTT-----GGC 423
|||:||| ||| ||||| :|||::: |||::: |||
Db 167 ProLysGlyTrpGlyPheSerIleAlaGlyGlyValGlyTy rGlnHisIleProGlyAsp 186
Qy 424 TGTGGCTCTTCATCTCCCACCTCATCAAGCGGTGAGGAGCACGCTCGGG---CTC 480
|||::: |||::: |||::: |||::: |||::: |||::: |||
Db 187 AsnSerIleAlaValThrLysIleIleGluGlyGlyAlaAspHisLysAspGlyArgLeu 206
Qy 481 CAGTAGGGACGAGATCGTCGGATCAATGATATTCATCTCTCTGTACCCTGAG 540
|||:|||:|||:|||:|||:|||:|||:|||
Db 207 GlnIleGlyLysIleLeuAlaValAlaAsnSerValGlyPheGluAspValMetHisGlu 226
Qy 541 GAGTGTCAACCTCATTCGA---ACCAAGAAAACTGTCTCATCAAGTGAAGACACATC 597
::: ::::: |||::: |||::: |||::: |||::: |||
Db 227 AspAlaValGlyAlaLeuLysAsnThrTy rAspValValHisLeuLysValAlaLysPro 246
Qy 598 GGCCTGATCCCGTGAAAGCTCT-----CCTGATGAGCCCCCTCACTTGG---CAG 645
|||::: |||::: |||::: |||::: |||::: |||
Db 247 SerAsnAlaIleLeuSerAspSerTy rAlaProAspLysThrThrSerTy rSerGln 266
Qy 646 TATGTGATCAGTTTGTGCGGAATCTGGGCGGTGCGAGGACGCTGGG----- 696
:|||:||| |||::: |||::: |||::: |||
Db 267 HisLeuAspLeuGluIleSerHisSer-----SerTy rLeuGlyMetAspTy r 282
Qy 697 -----TCCCTCGAAATCGGAA----- 714
|||::: |||::: |||::: |||::: |||::: |||

Db 283 ProThrAlaMetThrProThrAsnProArgArgTr y rSerProValAlaLysGlnLeuLeu 302

QY 715 -----AACAGGAGAGAGAGGCTTTCATCAGCCTGGTAGGCTCCCGA 756
Db 303 GlyGluGluAspIleProArgArgProArgIleValIleHisArg---GlyThrThr 321
QY 757 GGCCTTGCTGCAGCATTTCCAGCGGCCCATCCAGAGCCCTGGCATCTTTATCAGCCAT 816
Db 322 GlyLeuGlyPheAsnIle---ValGlyValGluAspGlyGluGlyIlePheIleSerTrp 340
QY 817 GTGAACCTGGCTCCTCTGCTGCTGAGGTGGA---TTGGAGATAGGGGACAGATTGTC 873
Db 341 IleLeuAlaGlyGlyProAlaAspLeuTyrglyGluLeuArgLysGlyAspGlnIleAla 360
QY 874 GAAGTCAATGGCGTCGATCTCTTAACCTGGATCAAGAGGCGTAAATGCTGTA 933
Db 361 SerValasnGlyValasLeuArgasnaspSerHisGluGlnAlaAlaIleAlaLeuGlu 380
QY 934 AATAGCCGAGCCTGACCATCTCCATTGTAGCTGCGCGGAGCTGTTTCATGACA 993
Db 381 AsnAla---GlyGlnThrValThrIlePheGlnTyrlsProGluGluTyrlsSerArg 399
QY 994 GACCGGAGCGCTGGCAGCGCGGCGAGCTGAGCTGCAGCGCAG-----1041
Db 400 GlyGluAlaLysIleHisAspLeuArgGluHisLeuMetAsnSerSerLeuGlySerGly 419
QY 1042 -----GAGCTT 1047
Db 420 IleAlaSerLeuArgSerAsnProLysArgLysPheTyrlleArgAlaLeuPheAspTrp 439
QY 1048 CTCATCAGAGAGCGG-----CTGGCGATG-----GAGTCC 1077
Db 440 LeuLysThrLysAspCysGlyPheLeuSerMetAlaLeuSerPheHisPheGlyAspVal 459
QY 1078 ACAAGATCCTCCAGGAGCAGGAGATGCGCGGCAAGAGAGAAAGAAATTCGCCAG 1137
Db 460 AsnHisValIleAspAlaSerAspGluGluGlnTrpGlnAlaArgValHisSerAsp 479
QY 1138 AAGCCAGCAGAGAA-----AATGAGATATCCGGAAGAGATGCAACAGATT 1185
Db 480 ArgGluThrAspAspIleGlyPheIleProThrLysArgArgValGluArgGluTrp 499
QY 1186 GTAGAGGAGAGAGAGATTTAAGAGCAATGGGAAGAGATGGGGCTCAAAGGAACAG 1245
Db 500 ValArgLeuLysAlaLysAspTrpGlySerTrpSerGlySerGlnGlyArgGluAspSer 519
QY 1246 CTACTCTTCCCTAAACCATCACT---GCTGAGGTACAC-----CCAGTACCCCTT 1293
Db 520 TyrLeuSerTyrgluThrValThrGlnMetAlaValHisTyrlleAlaArgProIleIleIle 539
QY 1294 CGCAAGCCAAAGTATGATCAGGAGTGGAACTGAGCTCGAGCCCGAGATGACCTGGAT 1353
Db 540 AspGlyProThrLysAspArgAlaAsnAspGluLeuLeuSerGluPheProAspLysPhe 559
QY 1354 GGAGGACGAGGAGGAGGAGCAGAGATTTCCGGAATATGAG---GAAGCTTTGAC 1410
Db 560 PheSerCysValProHisThrThrArgProGlyArgGluTyrglyIleAspGlyArgAsp 579
QY 1411 CCTACTCTATGTCACCCAGCAGATCATCGGGAAGATGTCGGCTCTCCATCCATC 1470
Db 580 HisHisPheValSerArgGluLysMetIle---LysAspIleGlnAlaHisLysPhe 598
QY 1471 AACAAGGAGGATCCTTAGACCTGGCCCTGGAAGCGGTGTGAGCTCCCGGAGGAG 1530
Db 599 IleLysAlaGlyGlnTyrlsSerHisLysLeuTyrlsSerValGlnSerValArgGlu 618
QY 1531 GTGGTC-----GTTTCTGCTGTGTATGAGCGG 1557
Db 619 ValMetGluGlnGlyLysHisCysIleLeuAsnSerAlaAsnAlaValArgArgLeu 638
QY 1558 GGAGCTGTGAGCGGCGATGCTGTCATGTG-----AAAGGGGACGAGATC 1602
Db 639 GlnGlnAlaHisLeuHisProIleAlaIlePheArgArgProArgSerLeuGluAsnVal 658

QY 1603 ATGCAATCAACGCAAGATTGTGACAGACTACACCTGGCTGAGGTGAGCTGCTGCC--- 1659
Db 659 LeuSerIleasnLysArgIleThrGluGluGlnThrArgLysAlaPheAspArgAlaThr 678
QY 1660 ---CTGCAAGAGGCTGGAATCAGGCGGGGAGCTGGATC 1695
Db 679 LysValGluGlnGluPheThrGluCysPheSerTrpIle 691
RESULT 11
US-09-045-632-14
; Sequence 14, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: HUGANIR, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-045-632-14
Alignment Scores:
Pred. No.: 1.76e-08 Length: 604
Score: 200.50 Matches: 127
Percent Similarity: 35.89% Conservative: 93
Best Local Similarity: 20.72% Mismatches: 222
Query Match: 4.87% Indels: 171
DB: 3 Gaps: 25
US-09-502-945-4 (1-2236) x US-09-045-632-14 (1-604)
QY 364 CTGACCTCTGACCCCGAAGGC-----CTGGCCTGAGTGTGCTGGCTGGAG 417
Db 3 ValGluLeuMetLysLysGluGlyThrThrLeuGlyCysThrValSerGlyGlyIleAsp 22
QY 418 TTTGGCTGTGGCTCTCTTCATCTCCACCTCATCAAGGCGGT---CAGGCACAGACGCTC 474
Db 23 LysaspGlyLysProArgValSerAsnLeuArgGlnGlyIleAlaAlaSerAsp 42
QY 475 GGGCTCCAGGTAGGGGACGAGATGTCGCGATCAATGGATATTCATCTCCTCTGTACC 534

Db 43 GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg 62
QY 535 CATGAGAGGTCTCAACCTCATTCGA---ACCAGAAACTGTCTCCATCAAGTGA 591
Db 63 HisAspGluIleLeuSerLeuGlyValGluValGluValGluValGlu 82
QY 592 CATCATGGCTGCTCCCTGAAAGCTCTCTGATGAGCCCTCACTTGGCAGTATG 651
Db 83 Tyr-----GluLeuProValSerIleGlnGlySerSerValMetPheArgThrVal 100
QY 652 GATCATGTTCTCGAATCTGGGCGTGGCGAGGAGCCTG-----GGTCCCTCGA 705
Db 101 GluValThrHisLysGluGlyAsnThrPheGlyPheValIleArgGlyAlaHis 120
QY 706 AATCGGAAACAAAGAGAGAGTCTTCATCAGCCTGGTAGGCTCCCGAGCCTGCG 765
Db 121 AspAspArgAsnLysSerArgProValIleThr----- 132
QY 766 TGCAGATTTCAGCGGCCCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAACCT 825
Db 133 CysValArgProGlyGlyProAspAspArgGluGlyThr-----IleLysPro 148
QY 826 GCCTCCCTGCTGCTGAGTGGATGGAGTAGGGACCATGTGCAAGTCAATGCG 885
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157
QY 886 GTCGACTTCTTAACCTGGATCACAGAGGCTGTAAATGTCTGAAA-----AAT 936
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177
QY 937 AGCCGAGCTGACCATC-----TCCATTGTAGCTGCAGCT 972
Db 178 GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer 197
QY 973 GCG----- 975
Db 198 GlyProLeuLeuValGluValAlaLysThrProGlyAlaSerLeuGlyValAlaLeuThr 217
QY 976 -----CGGAGCTGTTCATGACAGACCGGAGCGCTGGCAGAGCG 1017
Db 218 ThrSerValCysCysAsnLysGlnValIleValIleAspLysIleLysSerAlaSerIle 237
QY 1018 CGGACGCT-----GAGCTGCAGCGGAGGAGCTTCTC 1050
Db 238 AlaAspArgCysGlyAlaLeuHisValGlyAspHisIleLeuSerIleAspGlyThrSer 257
QY 1051 ATGCAGAGCGCTGGCGATGGAGTCCACAGATCTCTC-----CAGGAG 1095
Db 258 MetGluTyrCysThrLeuAlaGluAlaThrGlnPheLeuGlyAsnThrThrAspGlnVal 277
QY 1096 CAGCAGAGATGGAGCGCAAGAGAGAAAGAAATGGCCAGAGCGCAGAGAGAAAT 1155
Db 278 LysLeuGluIleLeuProHisGlnThrArgLeuAlaLeuLysGlyProAspHisVal 297
QY 1156 GAGAGATACCGAGAGAGATGGAACAGATTGTAGAGGAGAGAGAACTTTAAGAGCAA 1215
Db 298 LysIleGlnArgSerAspArgGln-----LeuPro 307
QY 1216 TGGNAGAGACTGGGCTCAAGGACAGCTACTCTTCCTAAACCATCACTGCTGAG 1275
Db 308 Trp---AspProTrpAlaSerSerGlnCysSerValHisThrAsnHisHisAsnPro 326
QY 1276 GTACACCCAGTACCTTCGCAAGCCAAAG-----TATGATCAGGAGTGGACCTGAG 1329
Db 327 HisHisProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsn 346
QY 1330 CTCGAGCCGCGAGTACCTGGATGGAGCGAGGAGGAGGAGGAGGAGGATTCGCG 1389
Db 347 SerProAlaMetValSerSerSerSerProThrSerMetSerAlaTyrSerLeuSer 366
QY 1390 AAATATGAGGAGGCTTTGACCCCTACTCTATGTTCT-----ACCCAGACAGCATCATG 1443
Db 367 SerLeuAsnMetGlyThrLeuProArgSerLeuTyrSerThrSerProArgGlyThrMet 386

QY 1444 GGAAGGATGTCCGCTCCTACGATCAAGAAG-----GAGGGATCCTTTAGACCTG 1494
Db 387 MetArg-----ArgArgLeuLysLysAspPheLysSerSerLeuSerLeu 402
QY 1495 GCC----- 1497
Db 403 AlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrGluValVal 422
QY 1498 -----CTGGAAGCGGTGTG----- 1512
Db 423 LeuThrAlaAspProValThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThr 442
QY 1513 -----GACTCCCCCATGGGAAGGTGCTTCTGCTGTGTATGACGGGGAGCT 1563
Db 443 GluThrLeuSerSerPro-----ProLeuIleSerTyrIleGluAlaAspSerPro 459
QY 1564 GCTGAGCGGCATGTGGCATTTGAAAGGGGACGAGATCATGGCAATCAACGGCAAGATT 1623
Db 460 AlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIlePro 479
QY 1624 GTGACAGACTACACCTGGCTGAGCTGACGCTGCCCTGCAGAGGCTTGAATCAGGGC 1683
Db 480 ThrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerSerIleThrSer 499
QY 1684 GGGAGCTGATC-----GACCTTGTGTGCTGCCCTCTGCCCTC----- 1719
Db 500 LysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerSerGlyThr 519
QY 1720 -----CCAAAGGAGTATGACGATGACGCTGACCTTCTTGTGCTGAAGTCC--- 1761
Db 520 PheHisValLysLeuProLysLysHisSerValGluLeuGlyIleThrIleSerSerPro 539
QY 1762 -----AAAAGGGGAAACCAAAATT 1779
Db 540 SerSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlySerValAla 559
QY 1780 CACCGGTAGAAACAGTACGCTCGGCCCGCCACCTCGTG 1818
Db 560 HisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572
RESULT 12
US-09-045-632-15
; Sequence 15, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: HUGANIR, Richard L.
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0; Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860

; REFERENCE/DOCKET NUMBER: 48147/1699-CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-523-3400

; TELEFAX: 617-523-6440

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 702 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-045-632-15

Alignment Scores:

Pred. No.: 1,91e-08 Length: 702

Score: 200.50 Matches: 127

Percent Similarity: 35.89% Conservative: 93

Best Local Similarity: 20.72% Mismatches: 222

Query Match: 4.87% Indels: 171

DB: 3 Gaps: 25

US-09-502-945-4 (1-2236) x US-09-045-632-15 (1-702)

```
QY 364 CTGGACGCTGCACCCGGAAGC-----CTGGCCCTGAGTGTGCGGTGGCTGGAG 417
Db 3 ValGluLeuMetLysLysGluGlyThrThrLeuGlyCysThrValSerGlyGlyIleAsp 22
QY 418 TTGGCTGTGGGCTTTCATCTCCACCTCATCAAGCGGT--CAGCGACACAGCGTC 474
Db 23 LysAspGlyLysProArgValSerAsnLeuArgGlnGlyIleAlaAlaArgSerAsp 42
QY 475 GGGCTCCAGGTAGGGAGCAGATCGTCGGATCAATGATATTCGATCTCCCTGTACC 534
Db 43 GlnLeuAspValGlyAspTrpIleLysAlaValAsnGlyLeuAsnLeuAlaLysPheArg 62
QY 535 CATGAGGAGGTATCAACCTCATTCGA---ACCAGAAACATGTGTCCATCAAGTGAGA 591
Db 63 HisAspGluIleLeuSerLeuLysAsnValGlyGluArgValValLeuGluValGlu 82
QY 592 CACATCGGCTGATCCCGCTGAAAGCTCTCTGTATGAGCCCTCACTTGGCAGTATGG 651
Db 83 Tyr-----GluLeuProValSerIleGlnGlySerValMetPheArgThrVal 100
QY 652 GATCAGTTTGTGGAATCTGGGGCGTGGCGAGGCGCTG-----GGCTCCCTGGA 705
Db 101 GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyAlaHis 120
QY 706 AATCGGGAACAAAGAGAGAGTCTTCATCAGCTCGTAGGCTCCCGAGCCTGGC 765
Db 121 AspAspArgAsnLysSerArgProValValIleThr----- 132
QY 766 TGACGATTTCCAGCGGCCCCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAACCT 825
Db 133 CysValArgProGlyLysProAspArgGluGlyThr-----IleLysPro 148
QY 826 GGCTCCCTGCTGCTGAGGTGGGATGGAGATAGGGACCATGATGTCGAAGTCAATGCG 885
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157
QY 886 GTCGCTTCTTAACCTGATCAACAGGAGGCTGTAATGTGCTGAAA-----AAT 936
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177
QY 937 AGCCGAGCGCTGACCATC-----TCCATTGTAGTCGAGCT 972
Db 178 GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer 197
QY 973 GGC----- 975
Db 198 GlyProLeuValGluValAlaLysThrProGlyAlaSerLeuGlyValAlaLeuThr 217
QY 976 -----CGGAGCTGTTCATGACAGACCGGGAGCGGTGGCAGAGCGG 1017
Db 560 HisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572
```

```
Db 218 ThrSerValCysCysAsnLysGlnValIleValIleAspLysLysIleLysSerAlaSerIle 237
QY 1018 CGGCAGCGT-----GAGCTGCAGCGCGCAGGAGCTTCTC 1050
Db 238 AlaAspArgCysGlyAlaLeuHisValGlyAspHisIleLeuSerIleAspGlyThrSer 257
QY 1051 ATGCAGAAAGCGGTGGCGATGGAGTCCCAACAAGATCCTC-----CAGGAG 1095
Db 258 MetGluTyrCysThrLeuAlaGluAlaThrGlnPheLeuGlyAsnThrThrAspGlnVal 277
QY 1096 CAGCAGGAGATGGAGCGGCAAGAGAAAGAAATATCCCAAGAGCAGCAGAGGAAAT 1155
Db 278 LysLeuGluIleLeuProHisHisGlnThrArgLeuAlaLeuLysGlyProAspHisVal 297
QY 1156 GAGAGATACCGGAGGAGATGGAAACAGATTCTAGAGGAGGAAGAGATTAAAGAACAA 1215
Db 298 LysIleGlnArgSerAspArgGln-----LeuPro 307
QY 1216 TGGGAAGAAGACTGGGGCTCAAGGAACAGACTACTTGGCTAAACCATCATCTGCTGAG 1275
Db 308 Trp---AspProTrpAlaSerSerGlnCysSerValHisThrAsnHisHisAsnPro 326
QY 1276 GTACACCCAGTACCCCTTCGCAAGCCAAAG-----TATGATCAGGGAGTGGAACTGAG 1329
Db 327 HisHisProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsn 346
QY 1330 CTCGAGGCCGCGAGATGACCTGGATGGAGGACGAGGAGGAGGAGGAGGAGGAGTTCGG 1389
Db 347 SerProAlaMetValSerSerSerSerProThrSerMetSerAlaTyrSerLeuSer 366
QY 1390 AAATATAGGAAAGCTTTGACCCCTACTCTATGTTTC-----ACCCAGAGCAGATCATG 1443
Db 367 SerLeuAsnMetGlyThrLeuProArgSerLeuTyrSerThrSerProArgGlyThrMet 386
QY 1444 GGGAGGATGTCGGCTCTACGCATCAAGAG-----GAGGGATCTCTAGACCTG 1494
Db 387 MetArg-----ArgArgLeuLysLysLysAspPheLysSerSerLeuSerLeu 402
QY 1495 GCC----- 1497
Db 403 AlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrThrGluValVal 422
QY 1498 -----CTGGAAGCGGTGTG----- 1512
Db 423 LeuThrAlaAspProValThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThr 442
QY 1513 -----GACTCCCGCATTTGGGAAGTGTCTGCTGTGTATGATGACGGGAGCT 1563
Db 443 GluThrLeuSerSerPro-----ProLeuIleSerTyrIleGluAlaAspSerPro 459
QY 1564 GCTGAGCGCGCATGTGGCATTTGTAAAGGGGACGAGATCATGGCAATCAACGGCAAGATT 1623
Db 460 AlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIlePro 479
QY 1624 GTGACAGACTACACCTCGCTGAGCTGACCTGCCCTGCGAGAGGCGCTGGAATCAGGGC 1683
Db 480 ThrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerSerIleThrSer 499
QY 1684 GGGGACTGGATC-----GACCTGTGTTGCTGCTGCCGCC----- 1719
Db 500 LysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerSerGlyThr 519
QY 1720 -----CCAAAGGAGTATGACATGATGATGATGATGATGATGATGATGATGATGAT 1761
Db 520 PheHisValLysLeuProLysLysHisSerValGluLeuGlyIleThrIleSerSerPro 539
QY 1762 -----AAAGGGGAACCAATTT 1779
Db 540 SerSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlySerValAla 559
QY 1780 CACGGTTAGGAACAGTACGCTCGGCCCCCACCCTCGTG 1818
Db 560 HisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572
```


RESULT 13

US-09-045-632-16
 ; Sequence 16, Application US/09045632
 ; Patent No. 6001575
 ; GENERAL INFORMATION:
 ; APPLICANT: Haganir, Richard L.
 ; APPLICANT: Dong, Hualing
 ; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
 ; TITLE OF INVENTION: GRIP-RELATED MOLECULES
 ; NUMBER OF SEQUENCES: 105
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
 ; STREET: 130 Water Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patencin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/045,632
 ; FILING DATE: 19-MAR-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/041,016
 ; FILING DATE: 19-MAR-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Corless, Peter F.
 ; REGISTRATION NUMBER: 33,860
 ; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-523-3400
 ; TELEFAX: 617-523-6440
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1018 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-045-632-16

Alignment Scores:

Pred. No.:	2,34e-08	Length:	1018
Score:	200.50	Matches:	127
Percent Similarity:	35.89%	Conservative:	93
Best Local Similarity:	20.72%	Mismatches:	222
Query Match:	4.87%	Indels:	171
DB:	3	Gaps:	25

US-09-502-945-4 (1-2236) x US-09-045-632-16 (1-1018)

QY	364	CTGCAGCGTCTGACCCCGAAGGC-----CTGGCGCTGAGTGTGGTGGTGGCTGGAG	417
Db	3	Va1gluLeuMetLysLysGluGlyThrThrLeuGlyCysThrValSerGlyGlyIleAsp	22
QY	418	TTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGCGGT---CAGGCAGACACGGTC	474
Db	23	LysAspGlyLysProArgValSerAsnLeuArgGlnGlyGlyIleAlaAlaArgSerAsp	42
QY	475	GGGCTCCAGTAGGGAGGAGATCGTCGGGATCAATGGATATTCCTCTCTCTGTACC	534
Db	43	GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg	62
QY	535	CATGAGAGGTATCAACCTCATTCGA---ACCAAGAAACTGTGTCATCAAGTGAGA	591
Db	63	HisAspGluIleIleSerLeuLeuLysAsnValGlyGluArgValValLeuGluValGlu	82
QY	592	CACATCGGCTGATCCCCGTGAAAGCTCTCTGTGATGAGGCCCTCACTTGGCAGTATGTG	651

Db	83	Tyr-----GluLeuProProValSerIleGlnGlySerSerValMetPheArgThrVal	100
QY	652	GATCAGTTTGTGTCGGAATCTGGGGCTGCGAGGACGAGCTG-----GGCTCCCTGGA	705
Db	101	GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyGlyAlaHis	120
QY	706	AATCGGAAACAAAGAGAGAAAGGTCTTCATCAGCTCGTAGGCTCCCGAGCCCTGGC	765
Db	121	AspAspArgAsnLysSerArgProValValIleThr-----	132
QY	766	TGCAGCATTTCCAGCGCCCATCCAGAGCCCTGGATCTTTATCAGCCATGTGAACCT	825
Db	133	CysValArgProGlyGlyProAspArgGluGlyThr-----IleLysPro	148
QY	826	GGCTCCCTGTCTCTGAGGTGGATTGGAGATAGGGGACGACAGATTGTCAAGTCAATGGC	885
Db	149	-----GlyAspArgLeuLeuSerValAspGly	157
QY	886	GTGACTTCTTAACCTGGATCAAGAGGCTGTAAATGTGCTGAAA-----AAT	936
Db	158	IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln	177
QY	937	AGCCGAGCGCTGACCATC-----TCCATTGTAGCTGCAGCT	972
Db	178	GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer	197
QY	973	GGC-----	975
Db	198	GlyProLeuLeuValGluValAlaLysThrProGlyAlaSerLeuGlyValAlaLeuThr	217
QY	976	-----CGGAGCTGTTTCATGACAGACCGGAGCGGCTGGCAGAGCGG	1017
Db	218	ThrSerValCysCysAsnLysGlnValIleValIleAspLysIleLysSerAlaSerIle	237
QY	1018	CGCAGCGT-----GAGCTGCAGCGGAGGAGTCTCTC	1050
Db	238	AlaAspArgCysGlyAlaLeuHisValGlyAspHisIleLeuSerIleAspGlyThrSer	257
QY	1051	ATGCAGAAGCGCTGGGATGGATCCCAACAGATCTC-----CAGGAG	1095
Db	258	MetGluTyrCysThrLeuAlaGluAlaThrGlnPheLeuGlyAsnThrThrAspGlnVal	277
QY	1096	CAGCAGGATGGCGGCAAGGAGAAAGAAATTCGCCAGAGGACGACGAGAAAT	1155
Db	278	LysLeuGluIleLeuProHisGlnThrArgLeuAlaLeuLysGlyProAspHisVal	297
QY	1156	GAGAGATACCGNAGGAGATGGACAGATCTAGAGGAGGAGAGAGAGTTTAAAGACAA	1215
Db	298	LysIleGlnArgSerAspArgGln-----LeuPro	307
QY	1216	TGGGAAGAAGACTGGGGCTCAAAGGAACAGTACTCTTCCTAAACCATCACTGCTGAG	1275
Db	308	Trp---AspProTrpAlaSerSerGlnCysSerValHisThrAsnHisHisAsnPro	326
QY	1276	GTACACCCAGTACCCCTTCGCAAGCCAAAG-----TATGATCAGGGAGTGAACCTGAG	1329
Db	327	HisHisProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsn	346
QY	1330	CTGAGCCCGCAGATGACCTGGATGGAGCGCAGGAGGAGGAGGAGGAGGATTTCGGG	1389
Db	347	SerProProAlaMetValSerSerSerSerProThrSerMetSerAlaTyrSerLeuSer	366
QY	1390	AAATATGAGAAGGCTTTGACCCCTACTCTATGTTC-----ACCCAGAGCAGATCATG	1443
Db	367	SerLeuAsnMetGlyThrLeuProArgSerLeuThrSerProArgGlyThrMet	386
QY	1444	GGGAAGGATGTCGGGCTCTACGATCACAAG-----GAGGGATCTCTAGACCTG	1494
Db	387	MetArg-----ArgArgLeuLysLysLysAspPheLysSerSerLeuSerLeu	402
QY	1495	GCC-----	1497


```
Db 403 AlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrThrGluValVal 422
QY 1498 -----CTGGAAGCGCTGTG----- 1512
Db 423 LeuThrAlaAspProValThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThr 442
QY 1513 -----GACTCCCGCATTTGCTGCTGTGTATGAGCGGGAGCT 1563
Db 443 GluThrLeuSerPro-----ProLeuIleSerThrIleGluAlaAspSerPro 459
QY 1564 GCTGAGCGGATGTCGATGTGAAGGAGGAGATGTCGCAATGCAACGCGCAAGATT 1623
Db 460 AlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIlePro 479
QY 1624 GTGACAGACTACCCCTGGCTGAGCTGAGCTGCCCTGCGAGAAGGCTGGAATCAGGCG 1683
Db 480 ThrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerIleThrSer 499
QY 1684 GGGGACTGGATC-----GACCTTGTGTTGCCCTGCCCTCCCTCC----- 1719
Db 500 LysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerSerGlyThr 519
QY 1720 -----CCAAAGGAGTATGACGATGAGCTGACCTTCTTGTGCTGAAGTCC--- 1761
Db 520 PheHisValLysLeuProLysLysHisSerValGluLeuGlyIleThrIleSerSerPro 539
QY 1762 -----AAAAGGGGAAACCAAAATT 1779
Db 540 SerSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlySerValAla 559
QY 1780 CACGCGTTAGGAACAGTAGCTGCGGCCCCACCTCGTG 1818
Db 560 HisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572

RESULT 14
US-09-045-632-32
; Sequence 32, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Huginar, Richard L.
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045.632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1061 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-045-632-32

Alignment Scores:
Pred. No.: 2,39e-08 Length: 1061
Score: 200.50 Matches: 127
Percent Similarity: 35.89% Conservative: 93
Best Local Similarity: 20.72% Mismatches: 222
Query Match: 4.87% Indels: 171
DB: 3 Gaps: 25

US-09-502-945-4 (1-2236) x US-09-045-632-32 (1-1061)
QY 364 CTGACCGCTGACACCCGGAAGC-----CTGGCGCTGAGTGTGCTGGTGCCTGGAG 417
Db 3 ValGluLeuMetLysLysGluGlyThrThrLeuGlyCysThrValSerGlyGlyIleAsp 22
QY 418 TTTGGCTGTGGCTCTTCTATCTCCACCTCATCAAGGCGGT---CAGGCAGACAGCGTC 474
Db 23 LysAspGlyLysProArgValSerAsnLeuArgGlnGlyIleAlaAlaArgSerAsp 42
QY 475 GGGCTCAGTAGGGAGAGATCGCCGGATCAATGGATATTCATCTCTCTCTGTACC 534
Db 43 GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg 62
QY 535 CATGAGGAGTCTACACCTCATTCGA---ACCAAGAAACTGTGTCATCAAGTGAGA 591
Db 63 HisAspGluIleIleSerLeuLeuLysAsnValGlyIleArgValValLeuGluValGlu 82
QY 592 CACATCGCCCTGATCCCGTGAAGGCTCTCTGATGAGCCCTCACTTGGCAGTATGTG 651
Db 83 Tyr-----GluLeuProProValSerIleGlnGlySerSerValMetPheArgThrVal 100
QY 652 GATCAGTTTGTGCGAATCTGGGGCGTGCAGAGCAGCGCTG-----GGCTCCCTGGA 705
Db 101 GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyAlaHis 120
QY 706 AATCGGGAAACAGAGAGAGGCTCTTCATCAGCCTGCTAGGCTCCCGAGGCTTGGC 765
Db 121 AspAspArgAsnLysSerArgProValIleThr----- 132
QY 766 TGCAGCATTTCCACGCGCCCTCCAGAGCCTGGCATTTTATCAGCCATGTGAACCT 825
Db 133 CysValArgProGlyGlyProAspAspArgGluGlyThr-----IleLysPro 148
QY 826 GGCTCCCTGTCTGCTGAGTGGGATTCGAGATAGGGGACCAGATTGTCGAGTCAATGGC 885
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157
QY 886 GTCGACTTCTTAACCTGGATCACAGGAGGCTGTAATGTGCTGAAA-----AAT 936
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177
QY 937 AGCGCGACCTGACCATC-----TCCATTGTAGCTGCAGCT 972
Db 178 GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer 197
QY 973 GGC----- 975
Db 198 GlyProLeuLeuValGluValAlaLysThrProGlyAlaSerLeuGlyValAlaLeuThr 217
QY 976 -----CGGAGCTGTTATGACACGCGGAGCGGCTGGCAGAGCGG 1017
Db 218 ThrSerValCysCysAsnLysGlnValIleValIleAspLysIleLysSerAlaSerIle 237
QY 1018 CGGCAGCGT-----CAGCTGCGAGCGGAGGCTTCTC 1050
Db 238 AlaAspArgCysGlyAlaLeuHisValGlyAspHisIleLeuSerIleAspGlyThrSer 257
QY 1051 ATGCAGAGCGGCTGGCGATGGAGTCCAAACAAGATCCTC-----CAGGAG 1095
```



```
Db 258 MetGluTyrCysThrLeuAlaGluAlaThrGlnPheLeuGlyAsnThrThrAspGlnVal 277
QY 1096 CAGCAGAGATGGCGCGCAAGAGGAGAAAGAAATGGCCAGAGGCGCAGAGGAAAT 1155
Db 278 LysLeuGluIleLeuProHisGlnThrArgLeuAlaLeuLysGlyProAspHisVal 297
QY 1156 GAGAGATACCGGAGGAGATGGAACAGATTTGTAGAGGAGGAGAGAGAGATTTAAGAGCAA 1215
Db 298 LysIleGlnArgSerAspArgGln
QY 1216 TGGGAAGAAGACTGGGGCTCAAGAGACAGCTACTCTTCCTAAACCATCACTGCTGAG 1275
Db 308 Trp---AspProTrpAlaSerSerGlnCysSerValHisThrAsnHisHisAsnPro 326
QY 1276 GTACACCCAGTACCCTTCGGAAGCCAAAG-----TATGATCAGGGAGTGAACCTGAG 1329
Db 327 HisHisProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsn 346
QY 1330 CTCGAGCCCGCAGATGACCTGGATGGAGCGCAGGAGGAGGAGGAGGAGGATTTCCGG 1389
Db 347 SerProAlaMetValSerSerSerSerProThrSerMetSerAlaTyrSerLeuSer 366
QY 1390 AAATATGAGGAGGCTTTGACCCCTACTCTATGTTC-----ACCCAGAGCAGATCATG 1443
Db 367 SerLeuAsnMetGlyThrLeuProArgSerLeuTyrSerThrSerProArgGlyThrMet 386
QY 1444 GGGAGGATGTCGGCTCCTACGATCAAGAAG-----GAGGGATCCTTACAGCTG 1494
Db 387 MetArg-----ArgArgLeuLysLysLysAspPheLysSerSerLeuSerLeu 402
QY 1495 GCC----- 1497
Db 403 AlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrThrGluValVal 422
QY 1498 -----CTGGAAGCGGTGTG----- 1512
Db 423 LeuThrAlaAspProValThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThr 442
QY 1513 -----GACTCCCACTGGGAAGGTGCTTCTGCTGTGTATGAGCGGGGAGCT 1563
Db 443 GluThrLeuSerSerPro-----ProLeuIleSerTyrIleGluAlaAspSerPro 459
QY 1564 GCTGAGCGGCTGTGGCTATTGAAAGGGAGCAGATCATGGCAATCAACGCGCAAGATT 1623
Db 460 AlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIlePro 479
QY 1624 GTACAGACTACACCTGGCTGAGGCTGACGCTGCCCTGCAGAAAGGCTGGAATCAGGGC 1683
Db 480 ThrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerSerIleThrSer 499
QY 1684 GGGACTGGATC-----GACCTTGTGGTGGCTGCTGCC----- 1719
Db 500 LysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerSerGlyThr 519
QY 1720 -----CCAAAGGAGTATCAGCATGAGCTGACCTTCTTCTGCTGAAGTCC--- 1761
Db 520 PheHisValLysLeuProLysHisSerValGluLeuGlyIleThrIleSerSerPro 539
QY 1762 -----AAAAGGGGAAACCAATTT 1779
Db 540 SerSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlySerValAla 559
QY 1780 CACCGGTTAGAAACAGTACTCGGCCCCCACTCTGTG 1818
Db 560 HisArgThrGlyThrLeuGluSerLeuGlyAspLysLysLeuLeu 572
```

RESULT 15

US-09-562-737-22

; Sequence 22, Application US/09562737

; Patent No. 6428967

; GENERAL INFORMATION:

; APPLICANT: Herz, JoschIm

```
; APPLICANT: Gotthardt, Michael
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: US/09562,737
; CURRENT APPLICATION NUMBER: US/09562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-22
```

```
Alignment Scores:
Pred. No.: 2,36e-08 Length: 724
Score: 199.50 Matches: 122
Percent Similarity: 34.08% Conservative: 76
Best Local Similarity: 21.00% Mismatches: 149
Query Match: 4.84% Indels: 235
DB: 4 Gaps: 24
```

US-09-502-945-4 (1-2236) x US-09-562-737-22 (1-724)

```
QY 247 AATGAACCCAGCGCTGCTCTCTGTTTGTATGCCATTCGCCCTGATC----- 294
Db 24 HisSerProAlaHisLeuProAsnPheAlaAsnSerProProValIleValAsnGlyAsp 43
QY 295 -----CCACTGAAGCACCAGGTGAA-----TAT 318
Db 44 ThrLeuGluAlaProGlyTyrGluHisGlnValAsnGlyThrGluGlyGluMetIleTyr 63
QY 319 GATCAGCTGACCCCGCGCTCCAGGAAGCTGAAGGAGGTGCTGGACCGCTCTGCAC 378
Db 64 GluGluIleThrLeuGluArgGlyLys----- 72
QY 379 CCGGAAGCGCTCGGCGTGTGAGTGTGCGTGT-----GGCTGGAGTTTGGC 423
Db 73 ---SerGlyLeuGlyPheSerIleAlaGlyLeuThrAspAsnProHisIleGlyAspAsp 91
QY 424 TGTGGGCTCTTCATCTCCACCTCATCAAGGCGGTGATGATTTCCATCTCCTCTATCCCATGAG 540
Db 92 MetSerIlePheIleThrLysIleIleProAsnGlyAlaAlaGlnAspGlyArgLeu 111
QY 481 CAGTAGGGGACGACGATCGTCCGATCAATGGATATTTCCATCTCCTCTATCCCATGAG 540
Db 112 GlnValAsnAspSerIleLeuPheValAsnArgValAspValArgGluValThrHisSer 131
QY 541 GAGTCAATCAACCTCATTCGAAACCAAGAAACTGTGTCCATCAAGTGACACATCGGC 600
Db 132 SerAlaValGluAlaLeuLysGluAlaGlyThrIle-----ValArg 145
QY 601 CTGATCCCGTGAAGAGCTCTCTCTGATGAGCCCTCATCTTGGCAGTATGGGATCAGTTT 660
Db 146 LeuTyrValMetArgArgValProPro----- 154
QY 661 GTGTGGAATCTGGGGCGTGGGAGGCGCTGGGCTCCCTCGGAAATCGGAAACAAG 720
Db 155 -----Ala 155
QY 721 GAGAAGAGGTCTTCATCAGCCTGGTA---GGCTCCGAGGCGCTTGGCTGCACATTTCC 777
Db 156 GluLysIleIleGluIleThrLeuIleLysGlyProLysGlyLeuGlyTyrSerIleAla 175
QY 778 AGGGCCCCCATCCAGAAG-----CTGGC-----ATCTTATACCCCATGTGAAA 822
Db 176 GlyGlyValGlyAsnGlnAlaIleProGlyAspAsnSerIleTyrValAspLysIleIle 195
QY 823 CCTGGCTCCCTGCTGCTGAGGTGGGA---TTGGAGATAGGGACGACGATGTCGAGTGC 879
Db 196 GluGlyGlyAlaAlaHisGluAspGlyArgLeuGlnIleGlyAspLysPheLeuAlaVal 215
```


Alignment Scores:
Pred. No.: 2,63e-93 Length: 268
Score: 1350.00 Matches: 262
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 32.77% Indels: 0
DB: 10 Gaps: 0

US-09-502-945-4 (1-2236) x US-09-739-907-114 (1-268)

Qy 964 GCTGAGCTGCGCGGAGCTGTTATGACAGCCGGGAGCGCTGGCGAGCGCGGCGAG 1023
Db 6 AlaAlaAlaGlyArgGluLeuPheMetThrAspArgGluArgLeuAlaGluAlaArgGln 25
Qy 1024 CQTGAGCTGCAGCGCAGAGGCTTCTCATGAGAACCGCGCTGGCGATGAGTCCACAG 1083
Db 26 ArgGluLeuGlnArgGlnGluLeuMetGlnLysArgGluLeuAlaMetGluSerAsnLys 45
Qy 1084 ATCTCTCCAGGACGACGAGATGAGCGGCAAGGAGAAAGAAATTTGCCCAGAGGCA 1143
Db 46 IleLeuGlnGlnGlnGlnGlnMetGluArgGlnArgLysGluIleAlaGlnLysAla 65
Qy 1144 GCAGAGGAAATGAGAGATACCGGAGGAGATGGAACAGATTGTAGAGGAGAGAGAG 1203
Db 66 AlaGluGluAsnGluArgTyrArgLysGluMetGluGlnIleValGluGluGluLys 85
Qy 1204 TTTAAGAACCAATGGGAAGAGACTGGGGCTCAAGAGGACACTACTCTTGCCTAAACC 1263
Db 86 PheLysLysGlnTrpGluAspTrpGlySerLysGluGlnLeuLeuProLysThr 105
Qy 1264 ATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATGATCAGGAGTGAA 1323
Db 106 IleThrAlaGluValHisProValProLeuArgLysProLysTyrAspGlnGlyValGlu 125
Qy 1324 CQTGAGCTGCAGCCGACAGATGACCTGGATGAGGACGAGGAGGAGGAGAGCAGGAT 1383
Db 126 ProGluLeuGluProAlaAspLeuAspGlyGlyThrGluGluGlnGlyGluGlnAsp 145
Qy 1384 TTCCGGAATATGAGGAAGGCTTTACCCCTACTCTATGTTACCCCGAGAGCATCATG 1443
Db 146 PheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGluGlnIleMet 165
Qy 1444 GGAAGAGATGTCGGCTCTCAGCATCAAGAAGGAGGATCCTTAGACTGGCCCTGGAA 1503
Db 166 GlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeuAspLeuAlaLeuGlu 185
Qy 1504 GCGGTGTGAGATCCCCATTTGGGAGGTGTCGTTCTGCTGTATGAGCGGGAGCT 1563
Db 186 GlyGlyValAspSerProIleGlyLysValValSerAlaValTyrGluArgGlyAla 205
Qy 1564 GCTGAGCGCATGGTGGCATTTCTGAAGGGGACGAGATCATGSCAATCAACGGCAAGATT 1623
Db 206 AlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLysIle 225
Qy 1624 GTGACAGCTACACCTGCTGAGGCTGACGCTGCCCTGCAGAAGCCCTGGAATCAGGCG 1683
Db 226 ValThrAspTyrThrLeuAlaGluAlaAspAlaAlaLeuGlnLysAlaTrpAsnGlnGly 245
Qy 1684 GGGGACTGATGACCTTGTGTTGCTGCTGCCCTGCCCGCCAAAGAGTATGACCATGAGCTG 1743
Db 246 GlyAspTrpIleAspLeuValValAlaValCysProProLysGluTyrAspAspGluLeu 265
Qy 1744 ACCTTC 1749
Db 266 ThrPhe 267

RESULT 2
US-09-925-299-778
; Sequence 778, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 778
LENGTH: 120
TYPE: PRT
ORGANISM: Homo sapiens

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-778

Alignment Scores:
Pred. No.: 1,26e-25 Length: 120
Score: 447.00 Matches: 92
Percent Similarity: 87.04% Conservative: 2
Best Local Similarity: 85.19% Mismatches: 4
Query Match: 10.85% Indels: 10
DB: 9 Gaps: 2

US-09-502-945-4 (1-2236) x US-09-925-299-778 (1-120)

Qy 1315 GGAGTGAACCTGAGCTCGAGCCGCGAGATGACCTGGATGGAGCAGGAGGACGGA 1374
Db 23 GlyLeuArgProValLeuGlnPro-----ArgGlnGly 33
Qy 1375 GAGCAGGATTCCCGAAATATGAGGAAGGCTTGACCCCTACTCTATGTACCCAGAG 1434
Db 34 ---GlnAspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGlu 52
Qy 1435 CAGATCATGGGAAGGATGTCGGCTCTAGCATCAAGAGGAGGATCCTTAGACCTG 1494
Db 53 GlnIleMetGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeuAspLeu 72
Qy 1495 GCCCTGGAAGCGGTGTGGACTCCCCCATTTGGGAAGTGTGCTGCTGTGTATGAG 1554
Db 73 AlaLeuGluGlyGlyValAspSer**IleGlyLysValValSerAlaValTyrGlu 92
Qy 1555 CGGGAGCTGCTGAGCGCATGTCGTCATGTAAGAGGCGAGATCATGCGCAATCAAC 1614
Db 93 ArgGlyAlaAlaGluArgHisGlyIleValLysGlyAspGluIleMetAlaIleAsn 112
Qy 1615 GGCAAGATTGTGACAGACTACACC 1638
Db 113 GlyLysIleValThrAspTyrThr 120

RESULT 3
US-09-925-299-778
; Sequence 778, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 778
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens

FEATURE:
; NAME/KEY: SITE
; LOCATION: (81)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-778

Alignment Scores:
Pred. No.: 1.26e-25 Length: 120
Score: 447.00 Matches: 92
Percent Similarity: 87.04% Conservative: 2
Best Local Similarity: 85.39% Mismatches: 4
Query Match: 10.85% Indels: 10
DB: 10 Gaps: 2

US-09-502-945-4 (1-2236) x US-09-925-299-778 (1-120)

QY 1315 GGAGTGGACCTGACCTGAGCCGCCAGATGACCTGGATGGAGGACGAGGAGCAGGGA 1374
||||| ||| |||:||||
Db 23 GlyLeuArgProValLeuGlnPro-----ArgGlnGly 33
QY 1375 GAGCAGGATTTCCGGAATATGAGGAAGCTTTGACCCCTACTCTATGTTCAACCCAGAG 1434
||||| ||| |||:||||
Db 34 ---GlnAspPheArgLysGluGlyPheAspProTyrSerMetPheThrProGlu 52
QY 1435 CAGATCATGGGAAGATGTCGGCTCTACGATCAAGAAGGAGGATCCTTAGACCTG 1494
||||| ||| |||:||||
Db 53 GlnIleMetGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeuAspLeu 72
QY 1495 GCCCTGGAGCGGTGTCGACTCCCATTTGGGAAGTGGTGGTTCGTGTGTATGAG 1554
||||| ||| |||:||||
Db 73 AlaLeuGluGlyLysValAspSer***IleGlyLysValValSerAlaValTyrGlu 92
QY 1555 CGGGAGCTGCTGAGCGCATGTGGCATTTGTGAAGGGGAGGATCATGCAATCAAC 1614
||||| ||| |||:||||
Db 93 ArgGlyAlaAlaGluArgHisGlyLysValValLysGlyAspGluIleMetAlaIleasn 112
QY 1615 GCGAAGATGTGACAGACTACACC 1638
||||| ||| |||:||||
Db 113 GlyLysIleValThrAspTyrThr 120

RESULT 4

US-09-739-907-60
; Sequence 60, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; PRIOR FILING DATE: 2000-12-20
; PRIOR FILING DATE: 1999-07-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
FEATURE:
; NAME/KEY: SITE
; LOCATION: (98)
; OTHER INFORMATION: Xaa equals stop translation
US-09-739-907-60

Alignment Scores:

Pred. No.: 2.08e-23 Length: 98
Score: 417.00 Matches: 88
Percent Similarity: 83.81% Conservative: 0
Best Local Similarity: 83.81% Mismatches: 0
Query Match: 10.12% Indels: 17
DB: 10 Gaps: 1

US-09-502-945-4 (1-2236) x US-09-739-907-60 (1-98)

QY 676 GCGCTGCGAGCAGCCTCGGCTCCCTCGAAATCGGAAACAAAGAGAAAGTCTTC 735
||||| ||| |||:||||
Db 10 GlyValArgGlySerLeuGlySerProGlyAsnArgGluAsnLysGluLysLysValphe 29
QY 736 ATCAGCCTGTAGGCTCCGAGGCTTGGCTGCAGCATTTCCAGCGGCCCATCCAGAAG 795
||||| ||| |||:||||
Db 30 IleSerLeuValGlySerArgGlyLeuGlyCysSerIleSerSerGlyProIleGlnLys 49
QY 796 CCTGCGCATCTTATCAGCCATGTGAACCTGGCTCCCTGCTGCTGAGTGGGATTCGAG 855
||||| ||| |||:||||
Db 50 ProGlyIlePheIleSerHisValLysProGlySerLeuSerAlaGluValGlyLeuGlu 69
QY 856 ATAGGAGCAGATTTGCGAAGTCAATGGGCTCGACTTCTTAACCTGGATCAACAAGGAG 915
||||| ||| |||:||||
Db 70 IleGlyAspGlnIleValGluValAsnGlyValAspPheSerAsnLeuAspHisLysGlu 89
QY 916 GCTGTAATGTGCTGAAATAATAGCCGCGAGCTGACCATCTCCATTGTAGCTGCAGCTGGC 975
||| ||| |||:||||
Db 90 -Leu-----GlnLeuAl 93
QY 976 CGGGAGCTGTTC 988
||||| ||| |||:||||
Db 93 aGlySerCysSer 97

RESULT 5

US-09-739-907-119
; Sequence 119, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; CURRENT FILING DATE: 2000-12-20
; PRIOR FILING DATE: 1999-07-07
; PRIOR FILING DATE: 1999-07-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 119
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-907-119

Alignment Scores:

Pred. No.: 2.21e-23 Length: 113
Score: 417.00 Matches: 88
Percent Similarity: 83.81% Conservative: 0
Best Local Similarity: 83.81% Mismatches: 0
Query Match: 10.12% Indels: 17
DB: 10 Gaps: 1

US-09-502-945-4 (1-2236) x US-09-739-907-119 (1-113)

QY 676 GCGTGGAGGAGCTGGCTCCCTCGAAATCGGAAACAAAGAGAAAGTCTTC 735
||||| ||| |||:||||
||| ||| |||:||||


```
Db 26 GlyValArgGlySerLeuGlySerProGlyAsnArgGluAsnLysGluLysLysValPhe 45
QY 736 ATGACCTGTAGCTCCCGAGGCTTGGCTGAGAGATTTCCAGCGGCCCATCCAGAG 795
Db 46 IleSerLeuValGlySerArgGlyLeuGlycSerIleSerGlyProIleGlnLys 65
QY 796 CCTGGCATCTTTATCAGCATGTGAACCTGGCTGCCCTGCTGCTGAGGTGGGATTGGAG 855
Db 66 ProGlyIlePheIleSerHisValLysProGlySerLeuSerAlaGluValGlyLeuGlu 85
QY 856 ATAGGGACACAGATTGCGAAGTCAATGGCGTGCAGCTTCTTAACCTGGATCACAAGAG 915
Db 86 IleGlyAspGlnIleValGluValAsnGlyValAspPheSerAsnLeuAspHisLysGlu 105
QY 916 GCTGTAATGTGCTGAAAAATAGCGCAGCGCTGACCATCTCCATTGTAGTGCAGGTGGC 975
Db 106 Leu-----GlnLeuAl 109
QY 976 CGGAGCTGTTCA 988
Db 109 aglyserCysSer 113
RESULT 6
US-09-739-907-111
; Sequence 111, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09739,907
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 111
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-907-111
Alignment Scores:
Pred. No.: 1,37e-22 Length: 95
Score: 406.00 Matches: 90
Percent Similarity: 60.13% Conservative: 2
Best Local Similarity: 58.82% Mismatches: 1
Query Match: 9.85% Indels: 60
DB: 10 Gaps: 1
US-09-502-945-4 (1-2236) x US-09-739-907-111 (1-95)
QY 1752 GCTGAAGTCCAAAGGGGAAACCAATTCACGCTTAGGAAACAGTACGCTCCGGCCCA 1811
Db 3 SerGluValGlnLysGlyLysProAsnSer-ProLeuGlyAsnSerGluLeuArgProHi 22
QY 1812 CTCCTGTGAACAAAGCCCTCGACAGCAGCTTGAGAGAGCCACATGACACACCATG 1871
Db 22 sLeuValAsnThrLysProArgThrSerLeuGluArgGlyHis----- 36
QY 1872 GCATCCTTGGGACCTGAATCTATCACCAGGAATCTCAAACCTCCCTTTGGCCCTGAACCA 1931
Db 36 ----- 36
QY 1932 GGGCCAGATGAAGAACAGCTCGGGCCACTTTTTTTGAAGGCCCAATGCTGGAGGAAGGGAGC 1991
```

```
Db 36 ----- 36
QY 1992 AGCCAGCCGTTTGGGAGAGATCTCAAGGATCCAGACTCTCATTCCTTCTCTGGCCCA 2051
Db 37 -----Thr-IleProPheLeuTyrPro 44
QY 2052 GTGAATTTGGTCTCTCCAGCTTTGGGGACTCCTTCTTGAACCCCTAATAAGACCCAC 2111
Db 44 erLuePheGlyLeuSerGlnLeuTyrGlyThrProSerLeuAsnProAsnLysThrProL 64
QY 2112 TGGAGTCTCTCTCTCCATCCCTCTCTCTGCCCCTCTGCTTAATTCGTCAGGATG 2171
Db 64 euLueSerLeuSerLeuHisProSerProLeuProSerAlaLeuIleAlaAlaArgIle 84
QY 2172 TCACTCCAAACCTTACTCTGAGCTCATTAATAAAA 2206
Db 84 alThrProAsnLeuThrLeuSerSerLeuIleLys 95
RESULT 7
US-10-078-090-151
; Sequence 151, Application US/10078090
; Publication No. US2003004815A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karra, Kalpana
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and
; FILE REFERENCE: DEX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 151
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-078-090-151
Alignment Scores:
Pred. No.: 7,91e-16 Length: 260
Score: 318.50 Matches: 85
Percent Similarity: 52.32% Conservative: 39
Best Local Similarity: 35.86% Mismatches: 84
Query Match: 7.73% Indels: 29
DB: 9 Gaps: 6
US-09-502-945-4 (1-2236) x US-10-078-090-151 (1-260)
QY 358 GTGCTGTGGACCGCTGCACCCGAGGCTGCGCTGAGTGTGCTGTGCTGGAG 417
Db 1 LeuArgArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleArgGlySerGlu 19
QY 418 TTGGCTGTGGCTTTCCTCTCCACCTCATCAAGCGGTGAGGACGACAGCGTGGG 477
Db 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
QY 478 CTCAGGTAGGGACGAGATCCCGCATCAATGATATTCATCTCTCTCTACCCAT 537
Db 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
QY 538 GAGAGGTGATCAACCTTCATTGCAACCAAGAAACTGTCTCCATCAAGGTGACACATC 597
Db 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
QY 598 GGCCTGATCCCGGTGAAAGACTCTCTGATGAGCCCTCTCTGCGCAGTATGTGGAT--- 654
```


Db	80	GlyArgIleProGlyGlyTyValThrAsnHisIleTyrThrTrp-----ValAspPro	97
QY	655	-----CAGTTGTGTCGCAATCTCGGGCGTGGAGGCAGCGCTGGC-----	696
Db	98	GlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg	117
QY	697	---TCCCTCGGAATCGGAAAC-----AAGGAGAAGAAG	729
Db	118	GlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLys	137
QY	730	GTCATCATCAGCGCTGGTAGGCTCCCGAGGCGTGGCTGCAGCAATTTCCAGCGGCCCATC	789
Db	138	ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu	157
QY	790	CAGAAGCGCTGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGCTCTCTGAGTGGGA	849
Db	158	TyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluGlyGlySerGly	177
QY	850	TTGAGATAGGGGACCAAGATGTGCAAGTCAATGGCGCTCGACTTCTCTAACTGGATCAC	909
Db	178	LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis	197
QY	910	AAGGAGGCTGPAATGTGCTGAAAAATAGCCGCGAGCTGACCATCTCCATTGTAGCTGCA	969
Db	198	AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThrVal-----	214
QY	970	GCTGCCGGGAGCTGTTTCATGACAGACCGGAGCGGCTGGCAGAGCGCGG	1020
Db	215	-----LysAspValGlyArgLeuProHisAlaArg	224

```

RESULT 8
US-09-951-402-3
; Sequence 3, Application US/09951402
; Patent No. US20020168752A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/951,402
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 09/306,998
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 60/084,740
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-951-402-3

Alignment Scores:
Pred. No.: 1,29e-11 Length: 2037
Score: 267.50 Matches: 142
Percent Similarity: 35.81% Conservative: 109
Best Local Similarity: 20.26% Mismatches: 223
Query Match: 6.49% Indels: 228
DB: Gaps: 26

US-09-502-945-4 (1-2236) x US-09-951-402-3 (1-2037)
QY 388 CTCGGCTGAGTGTGGTGGTGGCCCTGGAGTTGGCTGTGGGCTCTTCATCTCCACCTC 447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1018 LeuGlyMetThrValSerAlaAsnLysAsp---GlyLeuGlyMetIleValArgSerIle 1036
QY 448 ATCAAGGCGGTGACGGCAGACACGCTCGGG---CTCCAGGTAGGGGAGAGATCGTCGG 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1037 IleHisGlyGlyAlaIleSerArgAspGlyArgIleAlaIleGlyAspCysIleLeuSer 1056
QY 505 ATCAATGGATATTCATCTCTCTCTGTACCCATGAGGAGGTGTCATCAACCTCATTCGAA 564

```


Db 1370 GlyGlnIleLeuTyrGlyArgSerHisGlnAsnAlaSerSerIleIleLysCysAlaPro 1389
Qy 1393 -----TATGAGGAGGCTTTGAC 1410
Db 1390 SerLysValLysIleIlePheIleArgAsnLysAspAlaValAsnGlnMetAlaValCys 1409
Qy 1411 CCCTACTCTATGTCACCCCA-----GAGCAGATCATGGGAGGATGTC 1455
Db 1410 ProGlyAsnAlaValGluProLeuProSerAsnSerGluAsnLeuGlnAsnLysGluThr 1429
Qy 1456 CGGCTCTACGATCAAGAAGGAGTCCTTAGCTGGCCCTG----- 1500
Db 1430 GluProThrValThrSerAspAlaAlaValAspLeuSerSerPheLysAsnValGln 1449
Qy 1501 -----GAGGCGGTG-----GACTCCCCC 1521
Db 1450 HisLeuGluLeuProLysAspGlnGlyLeuGlyIleAlaIleSerGluGluAspThr 1469
Qy 1522 ATTGGAAGGTGGTCTGCTGCTGATGAGCGGGAGCTGCTGAGCGCATGGTGGC 1581
Db 1470 LeuSerGlyValIleIleLysSerLeuThrGluHisGlyValAlaAlaThrAspGlyArg 1489
Qy 1582 ATTGGAAGGGGAGGATGATGCAATCAAGCGCAAGATGTGACAGACTACACCTG 1641
Db 1490 LeuLysValGlyAspGlnIleLeuAlaValAspAspGluIleValValGlyTyrProIle 1509
Qy 1642 GCTGAGGCTGAGCGCTGCGCAGAAAGCGCTGGAATCAGGCGGGGACTGGATCGACCTT 1701
Db 1510 GluLysPheIleSerLeuLysThrAla----- 1519
Qy 1702 GTGGTGGCTGCTGCCCCCAAGGAGTATGACGATGAGCTGACCTTCTGCTGAA-GTC 1760
Db 1520 -----LysMetThrValLysLeuThrIle 1527
Qy 1761 CAAGAAGGGAACCAATTCACCGCTTAGGAAACAGTAGTCCGCGGCCACCTCGTGA 1820
Db 1528 HisAlaGluAsnProAspSerGlnAlaValProSerAlaAlaGlyAlaAlaSerGlyGlu 1547
Qy 1821 CACAAAGCTCGGACCGCTTGAGAGGCGCACATGACACACAGATGCGATCCCTTG 1880
Db 1548 LysLysAsnSerSerGlnSerLeuMetValPro-----GlnSerGlySerPro 1563
Qy 1881 GGACCTGAATCTATC-----ACCAGGAATCTCAACTCC-----TTTGGCCCTGAA 1928
Db 1564 GluProGluSerIleArgAsnThrSerArgSerThrProAlaIlePheAlaSerAsp 1583
Qy 1929 CCA----- 1931
Db 1584 ProAlaThrCysProIleIleProGlyCysGluThrThrIleGluIleSerLysGlyArg 1603
Qy 1932 -----GGCCAGATAAGGAACAGCTCGGGCCACTTTTGA 1967
Db 1604 ThrGlyLeuGlyLeuSerIleValGlyGlySerAspThrLeuLeuGlyAlaIleIleIle 1623
Qy 1968 AGGCCAATGTGAGGAAGGAGCAGCCAGCGCTTTGGGAGAGATCTCAAGGATCCAGA 2027
Db 1624 HisGluValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaGlyaspGln 1643
Qy 2028 CTC 2030
Db 1644 Ile 1644

RESULT 9

US-09-951-401-3
; Sequence 3, Application US/09951401
; Patent No. US20020115104A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Ravitigian, Sean V.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/951,401
; CURRENT FILING DATE: 2001-09-14

; PRIOR APPLICATION NUMBER: US 09/306,998
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 60/084,740
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-951-401-3

Alignment Scores: 1.29e-11 Length: 2037
Pred. No.: 267.50 Matches: 142
Score: 35.81% Conservative: 109
Best Local Similarity: 20.26% Mismatches: 223
Query Match: 6.49% Indels: 228
DB: 10 Gaps: 26

US-09-502-945-4 (1-2236) x US-09-951-401-3 (1-2037)

Qy 388 CTGGGCTGAGTGTGCGTGGCTGGCTGCTTCATCTCCACCTC 447
Db 1018 LeuGlyMetThrValSerAlaAsnLysAsp---GlyLeuGlyMetIleValArgSerIle 1036
Qy 448 ATCAAGGCGGTGACGAGCAGCGCTCGG---CTCCAGGTAGGAGCAGATCGTCCGG 504
Db 1037 IleHisGlyGlyAlaIleSerArgAspGlyArgIleAlaIleGlyAspCysIleLeuSer 1056
Qy 505 ATCAATGGATATTCATCTCTCTGTACCATGAGGAGTGTATCAACCTCATCTCGAACC 564
Db 1057 IleAsnGluGluSerThrIleSerValThrAsnAlaAlaArgAlaMetLeuArg--- 1075
Qy 565 AAGAAACTGTGTCCATCAAGTACAGACATCGGCTGATCCCGTGAAGAAAGCTCTCCT 624
Db 1076 -----ArgHis---SerLeuIle-----GlyPro 1082
Qy 625 GATGAGCCCTCACTTGG-----CAGTATGTGATCAGTTTGTGTCGGAATCTGGG 675
Db 1083 AspIleLysIleThrValProAlaGluHisLeuGluGluPheLysIleSerLeuGly 1102
Qy 676 GGCTGGGAGCAGCCCTG----- 693
Db 1103 GlnGlnSerGlyArgValMetAlaLeuAspIlePheSerSerTyrThrGlyArgAspIle 1122
Qy 694 -----GGCTCCCTGGAATCGGGAAACAAGAG----- 723
Db 1123 ProGluLeuProGluArgGluGluGlyGluSerGluLeuGlnAsnThrAla 1142
Qy 724 -----AAGAAGTC 732
Db 1143 TyrSerAsnTrpAsnGlnProArgArgValGluLeuTrpArgGluProSerLysSerLeu 1162
Qy 733 TTCTACAGCTGGTAGCTCCCGAGGCTTGGTGTGACGATTTCCAGCGGCCCATCCAG 792
Db 1163 GlyIleSerIleValGlyArgGlyMetGlySerArgLeuSerAsnGlyGluValMet 1182
Qy 793 AAGCCTGGCATCTTTATCAGCCATGTGAACCTGCTCCCTGCTGCTGAGTGGGA--- 849
Db 1183 Arg---GlyIlePheIleLysHisValLeuGluAspSerProAlaGlyLysAsnGlyThr 1201
Qy 850 TTGGAGATAGGGACCATGTCGAAGTCAATGCGCTGACCTTCTTAACCTGGATCAC 909
Db 1202 LeuLysProGlyAspArgIleValGluValAspGlyMetAspLeuArgAspAlaSerHis 1221
Qy 910 AAGGAGGCTGTAATGTGCTGAAATAAGCGCAGCTGACCATCTCCATGTAGCTGCA 969
Db 1222 GluGlnAlaValGluAlaIleArgLysAlaGlyAsnProValPheMetValGlnSer 1241
Qy 970 CTGCGCGGGAGCTGTTTCATACAGACGCGGAGCGGCTGGCAGAGCGCGCGCTGAG 1029
Db 1242 -----IleIleAsnArgProArgAlaProSerGlnSerGluSerGlu 1255

QY 1030 CTGACGGCGCAGGAGCTTCTCATGCAGAGCGGCTGGCGATGGAGTCCCAACAGATCCTC 1089
Db : : : : :
QY 1256 ProGluLysAlaProLeuCys----- 1262
QY 1090 CAGGACGACGAGGATGGAGCGGCAAGAGAAATTTGCCAGAGGACGACGAG 1149
Db : : : : :
QY 1263 -----SerValProProProSerAlaPheAlaGluMetGlySerAsp 1278
QY 1150 GAAATGAGACATCCGGAAGGAGATGGACACAGATTGTAGAGGAGGAAGAGATTTTAA 1209
Db : : : : :
QY 1279 HisThrGlnSerSerAlaSerLysIleSerGlnAspValAspLysGluAspGluPheGly 1298
QY 1210 AAGCAATGGGAA-----CAAGACTGGGCTCAAGGACACGACTACTCTTGCTTAAA 1260
Db : : : : :
QY 1299 TyrSerTrpLysAsnIleArgGluArgTyrGly----- 1309
QY 1261 ACCATCACTGCTGAGTACACCCAGTACCTCCCTTCGCAAG----- 1299
Db : : : : :
QY 1310 ThrLeuThrGlyGluLeuHisMetIleGluLeuGlyLysGlyHisSerGlyLeuGlyLeu 1329
QY 1300 -----CCA 1302
Db : : : : :
QY 1330 SerLeuAlaGlyAsnLysAspArgSerArgMetSerValPheIleValGlyIleAspPro 1349
QY 1303 AAGTATGATCAGGAGTGGACCTGAGCTCGAGCCGCGAGATGACCTG-----GAT 1353
Db : : : : :
QY 1350 AsnGlyAlaAlaGlyLysAspGlyArgLeuGlnIleAlaAspGluLeuGluIleAsn 1369
QY 1354 GGA-----GGCAGGAGGAGGAGGAGGAGGAGGATTTCCGGAAA----- 1392
Db : : : : :
QY 1370 GlyGlnIleLeuTyrGlyArgSerHisGlnAsnAlaSerSerIleIleLysCysAlaPro 1389
QY 1393 -----TATGAGGAAGGCTTTGAC 1410
Db : : : : :
QY 1390 SerLysValLysIleIlePheIleArgAsnLysAspAlaValAsnGlnMetAlaValCys 1409
QY 1411 CCCACTCTATGTTCCACCCCA-----GAGCAGATCATCGGGAAGAGATGTC 1455
Db : : : : :
QY 1410 ProGlyAsnAlaValGluProLeuProSerAsnSerGluAsnLeuGlnAsnLysGluThr 1429
QY 1456 CGGCTCTACGCATCAAGAGGAGGAGGATCTTAGACCTGGCGCTG----- 1500
Db : : : : :
QY 1430 GluProThrValThrSerAspAlaAlaValAspLeuSerPheLysAsnValGln 1449
QY 1501 -----GAGGCGGTGTG-----GACTCCCCC 1521
Db : : : : :
QY 1450 HisLeuGluLeuProLysAspGlnGlyLeuGlyIleAlaIleSerGluGluAspThr 1469
QY 1522 ATGGGAGGTGGTGGTTCCTGCTGTGTATGCGGGGAGCTGTGAGCGCATGGTGGC 1581
Db : : : : :
QY 1470 LeuSerGlyValIleIleLysSerLeuThrGluHisGlyValAlaAlaThrAspGlyArg 1489
QY 1582 ATTGTGAAAGGGGACGAGATCATGGCAATCAACGCAAGATTGTGACAGACTACACCTG 1641
Db : : : : :
QY 1490 LeuLysValGlyAspGlnIleLeuAlaValAspAspGluIleValGlyTyrProIle 1509
QY 1642 GCTGAGGCTGAGTGGCTGCTGCAGAGGCGCTGGAATCAGGCGGGGAGTGCATGCACCT 1701
Db : : : : :
QY 1510 GluLysPheIleSerLeuLeuLysThrAla----- 1519
QY 1702 GTGGTTGCGCTGCTGCCCCCAAGAGGATGATCAGATGAGCTGACCTTCTTCTGAA-GTC 1760
Db : : : : :
QY 1761 CAAAGGGGAAACCAATTACGGCTTAGGAAACAGTGCAGCTCGCGCCACCTCGTCAA 1820
Db : : : : :
QY 1528 HisAlaGluAsnProAspSerGlnAlaValProSerAlaAlaGlyAlaAlaSerGlyGlu 1547
QY 1821 CACAAGCCTCGGACCGCTTGTAGAGAGGCGCATGACACACACGATGGCATCTCTTGT 1880
Db : : : : :
QY 1548 LysLysAsnSerSerGlnSerLeuMetValPro-----GlnSerGlySerPro 1563

QY 1881 GGACCTGAATCTATC-----ACCCAGGAATCTCAAACTCCC-----TTTGGCCCTGAA 1928
Db : : : : :
QY 1564 GluProGluSerIleArgAsnThrSerArgSerThrProAlaIlePheAlaSerAsp 1583
QY 1929 CCA----- 1931
Db : : : : :
QY 1584 ProAlaThrCysProIleIleProGlyCysGluThrThrIleGluIleSerLysGlyArg 1603
QY 1932 -----GGCCAGATAAGGAACAGCTCGGGCCACTTTTTTGA 1967
Db : : : : :
QY 1604 ThrGlyLeuGlyLeuSerIleValGlyGlySerAspThrLeuLeuGlyAlaIleIle 1623
QY 1968 AGCCATGTGGAGGAAGGAGGAGCAGCAGCCGTTTGGGAGAGATCTCAAGATCCAGA 2027
Db : : : : :
QY 1624 HisGluValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaGlyAspGln 1643
QY 2028 CTC 2030
Db : : : : :
RESULT 10
US-09-922-101-3
; Sequence 3, Application US/09922101
; Patent No. US20020146711A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/922,101
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/306,998
; PRIOR FILING DATE: 1999-05-07
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-101-3
Alignment Scores:
Pred. No.: 1,29e-11 Length: 2037
Score: 267.50 Matches: 142
Percent Similarity: 35.81% Conservative: 109
Best Local Similarity: 20.26% Mismatches: 223
Query Match: 6.49% Indels: 228
DB: 10 Gaps: 26
US-09-502-945-4 (1-2236) x US-09-922-101-3 (1-2037)
QY 388 CTCGGCTGAGTGTGCTGGCTGGAGTTGGCTGTGGCTCTTCACTCCACCTC 447
Db : : : : :
QY 1018 LeuGlyMetThrValSerAlaAsnLysAsp--GlyLeuGlyMetIleValArgSerIle 1036
QY 448 ATCAAGGCGGTGAGGACAGCAGCGTGGG---CTCCAGGTAGGGACGAGATGTCGCGG 504
Db : : : : :
QY 1037 IleHisGlyGlyAlaIleSerArgAspGlyArgIleAlaIleGlyAspCysIleLeuSer 1056
QY 505 ATCAATGGATATCCATCTCTCTGTACCCATCAGGAGGTCATCAACCTCATTCGNAACC 564
Db : : : : :
QY 1057 IleAsnGluSerThrIleSerValThrAsnAlaAlaArgAlaMetLeuArg--- 1075
QY 565 AAGAAACTGTGTCCATCAAGTGAGACATCGGCTGATCCCGGTGAAAGCTCTCCT 624
Db : : : : :
QY 1076 -----ArgHis---SerLeulle-----GlyPro 1082
QY 625 GATGAGCCCTCACTTGG-----CAGTATGTGATCAGTTGTGTGCGAATCTGG 675
Db : : : : :
QY 1083 AspIleLysIleThrTyrValProAlaGluHisLeuGluGluPheLysIleSerLeuGly 1102
QY 676 GCGTGGCAGGACCCCTG----- 693
Db : : : : :


```
Db 1103 GlnGlnSerGlyArgValMetAlaLeuAspIlePheSerSerTyrThrGlyArgAspIle 1122
Qy 694 -----GGCTCCCTCGGAATCGGGAACAAAGGAG----- 723
Db 1123 ProGluLeuProGluArgGluGluGlyGluGluSerGluGlnAsnThrAla 1142
Qy 724 -----AAGAAGTGC 732
Db 1143 TyrSerAsnTrpAsnGlnProArgValGluLeuTrpArgGluProSerLysSerLeu 1162
Qy 733 TTCATCAGCCCTGGTAGCTCCGAGGCTTGGCTGCACATTTCCAGCGGCCCATCCAG 792
Db 1163 GlyIleSerIleValGlyArgGlyMetGlySerArgLeuSerAsnGlyGluValMet 1182
Qy 793 AAGCCTGGCATCTTATCAGCCCTCGAAACCTGGCTCGCTGCTGAGTGGGA--- 849
Db 1183 Arg---GlyIlePheIleLysHisValLeuGluAspSerProAlaGlyLysAsnGlyThr 1201
Qy 850 TTGGAGATAGGGACGACGATTGCTGAAGTCAATGGCGTGCATCTCTTAACCTGGATCAC 909
Db 1202 LeuLysProGlyAspArgIleValGluValAspGlyMetAspLeuArgAspAlaSerHis 1221
Qy 910 AAGGAGGCTGTAATGCTGTAAGAAATACCGCCAGCCTGACCATCTCCATTGCTAGCTGCA 969
Db 1222 GluGlnAlaValGluAlaIleArgLysAlaGlyAsnProValPheMetValGlnSer 1241
Qy 970 GCTGGCGGGAGCTGTTCATACAGACGCGGAGCGGCTGGCAGCGCGCGAGCCTGAG 1029
Db 1242 -----IleIleAsnArgProArgAlaProSerGlnSerGluSerGlu 1255
Qy 1030 CTGACGCGCAGAGCTTCTCATGAGAGCGGCTGGCGATGGATGCCAACAGATCCTC 1089
Db 1256 ProGluLysAlaProLeuGlyCys----- 1262
Qy 1090 CAGGAGCAGCAGAGATGAGCGGCAAGGAGAAAGAAATGCCAGAGGCCAGCAGAG 1149
Db 1263 -----SerValProProProProSerAlaPheAlaGluMetGlySerAsp 1278
Qy 1150 GAAATATGAGATACCGGAAGAGATGGAACAGATTGTAGAGGAGGAAGATTAAAG 1209
Db 1279 HisThrGlnSerSerAlaSerLysIleSerGlnAspValAspLysGluAspIleGly 1298
Qy 1210 AAGCAATGGGA-----GAAGACTGGGCTCAAGGAACAGCTACTCTTGCTTAA 1260
Db 1299 TyrSerTrpLysAsnIleArgGluArgTyrGly----- 1309
Qy 1261 ACCATCACTGCTGAGTACACCCAGTACCCCTCGCAAG----- 1299
Db 1310 ThrLeuThrGlyGluLeuHisMetIleGluLeuGluLysGlyHisSerGlyLeuGlyLeu 1329
Qy 1300 -----CCA 1302
Db 1330 SerLeuAlaGlyAsnLysAspArgSerArgMetSerValPheIleValGlyLeuAspPro 1349
Qy 1303 AAGTATGATCAGGAGTGAACCTGAGCTCGAGCCCGCAGATGACCTG-----GAT 1353
Db 1350 AsnGlyAlaAlaGlyLysAspGlyArgLeuGlnIleAlaAspLeuLeuGluIleAsn 1369
Qy 1354 GGA-----GGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1392
Db 1370 GlyGlnIleLeuTyrGlyArgSerHisGlnAsnAlaSerSerIleIleLysCysAlaPro 1389
Qy 1393 -----TATCAGGAAGGCTTTGAC 1410
Db 1390 SerLysValLysIleIlePheIleArgAsnLysAspAlaValAsnGlnMetAlaValCys 1409
Qy 1411 CCTACTCTATGTTACCCCA-----GACAGATCATGGGGAAGGATGTC 1455
Db 1410 ProGlyAsnAlaValGluProLeuProSerAsnSerGluAsnLeuGlnAsnLysGluThr 1429
Qy 1456 CGGCTCTAGCATCAAGAGGAGGATCTTAGACCTGGCCCTG----- 1500
Db 1430 GluProThrValThrThrSerAspAlaAlaValAspLeuSerSerPheLysAsnValGln 1449
```

```
Qy 1501 -----GAAGCGGTGTG-----GACTCCCC 1521
Db 1450 HisLeuGluLeuProLysAspGlnGlyLeuGlyIleAlaIleSerGluGluAspThr 1469
Qy 1522 ATTGGGAAGTGTCTGCTGCTGCTATACGCGGGAGCTGCTGAGCGGCATGGTGGC 1581
Db 1470 LeuSerGlyValIleIleLysSerLeuThrGluHisGlyValAlaAlaThrAspGlyArg 1489
Qy 1582 ATTGTGAAGGGGAGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTG 1641
Db 1490 LeuLysValGlyAspGlnIleLeuAlaValAspAspGluIleValGlyTyrProIle 1509
Qy 1642 GCTGAGCTGACGCTGCCCTGCACAGGCGCTGGAATCAGGCGGGGAGTGGATCGACCTT 1701
Db 1510 GluLysPheIleSerLeuLysThrAla----- 1519
Qy 1702 GTGTTGCGGTCTCCCGCCCAAGAGGAGTATGACGATGAGCTGACCTTCTTGTCTGAA-GTC 1760
Db 1520 -----LysMetThrValLysLeuThrIle 1527
Qy 1761 CAAAGGGGAACCAAAATTCACGCGTTAGGAACAGTGCAGCGCCACCTCGTGAA 1820
Db 1528 HisAlaGluAsnProAspSerGlnAlaValProSerAlaAlaGlyAlaAlaSerGlyGlu 1547
Qy 1821 CACAAAGCCTCGGACCGCTTGAGAGAGGCCACATCACACACAGATGGCATCCTTG 1880
Db 1548 LysLysAsnSerSerGlnSerLeuMetValPro-----GlnSerGlySerPro 1563
Qy 1881 GGACTCAATCTATC-----ACCAGGAATCTCAACTCCC-----TTTGGCCCTGAA 1928
Db 1564 GluProGluSerIleArgAsnThrSerArgSerThrProAlaIlePheAlaSerAsp 1583
Qy 1929 CCA----- 1931
Db 1584 ProAlaThrCysProIleIleProGlyCysGluThrThrIleGluIleSerLysGlyArg 1603
Qy 1932 -----GGCCAGATAAGGAACAGCTCGGCGCCACTTTTGTGA 1967
Db 1604 ThrGlyLeuGlyLeuSerIleValGlyGlySerAspThrLeuLeuGlyAlaIleIle 1623
Qy 1968 AGGCAATGTGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2027
Db 1624 HisGluValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaGlyAspGln 1643
Qy 2028 CTC 2030
Db 1644 Ile 1644
```

RESULT 11

```
US-09-919-497-59
; Sequence 59, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59
; LENGTH: 767
; TYPE: PRT
; ORGANISM: Homo sapiens
us-09-919-497-59
```

```
Alignment Scores:
Pred. No.: 1,7e-11 Length: 767
Score: 263.50 Matches: 125
Percent Similarity: 40.45% Conservative: 93
```



```

Best Local Similarity: 23.19%      Mismatches: 206
Query Match: 6.40%               Indels: 115
DB: 10                           Gaps: 24

US-09-502-945-4 (1-2236) x US-09-919-497-59 (1-767)

QY 340 TCACGAGAGCTGAAGAGGTGCTCTGGACCGTCTGCACCCCAAGGCGCTGCCTGAGT 399
Dy 198 AlslulysValMetGluilleLysLeuLys---GlyProlysGlyLeuGlyPheSer 216
QY 400 GTCGCTGGTGGC-----CTGGAGTTTGGCTGTGGCTCTCACTCCAC 444
Dy 217 IleAlaGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleYrValThrLys 236
QY 445 CTCATCAAGGCGGTGAGGAGACAGCGTGGG---CTCAGTAGGGGAGGATCGTC 501
Dy 237 IleleGluGlyAlaAlaHisLysAspGlyArgLeuGlnleGlyAspLysIleLeu 256
QY 502 CGGATCAATGGATATTCATCTCTCTCTGACCATGAGGAGTCACTCACTTCGA 561
Dy 257 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaAlaLys 276
QY 562 ---ACCAAGAAACTGTGTCATCAAGTG-----AGACATCGGCTGATCCCC 609
Dy 277 AsnThrTyraAspValValTyLeuLysValAlaLysProSerAsnAlaTyLeuSerAsp 296
QY 610 GTGAAAGCTCTCTGATGAGCCCTCCTCTG---CAGTATGTGATCAGTTTGTGTCG 666
Dy 297 SerTyraAlaProAspIleThrThrSerTyraSerGlnHisLeuAspAsnGluIleSer 316
QY 667 GAATCTGGGGCGTGGAGGACCCCTGGC----- 696
Dy 317 HisSer-----SerTyraLeuGlyThrAspTyraProThrAlaMetThrProThr 332
QY 697 -----TCCCTCGAAATCGGGA-----AACAG 720
Dy 333 SerProArgArgTyraSerProValAlaLysAspLeuLeuGlyGluAspIleProArg 352
QY 721 GAGAAGAGTCTTCATGACCGCTGGTAGGCTCCCGAGGCTGGTGCAGCATTTCCAGC 780
Dy 353 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 372
QY 781 GGCCCCATCCAGAGCGTGCATCTTTATCAGCCATGTG-----AACTGCTGCC 831
Dy 373 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 391
QY 832 CTGCTGCTGAGTGGGATGGAGTAGGGACAGATTGTGGAATGCTCAATGCGCTCGAC 891
Dy 392 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 409
QY 892 TTCTCTAACCTGGATCACAGAGGCTGTAAATGTCTGCTGAAATAGCCGCGAGCTGACC 951
Dy 410 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 428
QY 952 ATCTTCATGTAGTCCAGTCCGCGGAGCTTCTATCATCAGACCGGCGGCTGGCA 1011
Dy 429 ValThrIleIleAlaGln-----TyrLysProGluTyraSerArgPhe 443
QY 1012 GAGCGCGGCGAGCTGAGTGCAGCGGAGGAGCTTCTCATGAGAAGCGGTGGGATG 1071
Dy 444 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 461
QY 1072 GAGTCCACAGATCTCCAGGAGCAGCAGGAG-----ATGGAG 1110
Dy 462 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyraIleArgAlaLeuPheAsp 481
QY 1111 CGGCAAGAGCAAAAGAAATTCGCCAGAGGAGGAGCAAGAAAGAGATACCGGAAG 1170
Dy 482 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 501
QY 1171 GAGATGGAACAGATCTAGAGGAGGAGGAGAGAGTTT----- 1206
Dy 502 ValLeuHisValIleAspAlaSerAspGluGluTrpTrpGlnAlaArgValHisSer 521

```

```

QY 1207 -----AAGAGCAA 1215
Dy 522 AspSerGluThrAspAspIleGlyPheIleProSerLysArgValGluArgGlu 541
QY 1216 TGGGA-----GAAGCTGGGCTCA-----AAGAA 1242
Dy 542 TrpSerArgLeuLysAlaLysAspTrpGlySerSerSerGlySerGlnGlyArgGluAsp 561
QY 1243 CAGCTACTCTTCCTTAAACCATCACT---GCTGAGGTACAC-----CCAGTACC 1290
Dy 562 SerValLeuSerTyraGluThrValThrGlnMetGluValHisTyraAlaArgProIleIle 581
QY 1291 CTTCCCAAGCCAAAGTATGATCAGGAGTGGAACTCGAGCTCGAGCCGAGATCCTG 1350
Dy 582 IleLeuGlyProThrTyraAspArgAlaAsnAspLeuLeuSerGluPheProAspLys 601
QY 1351 GATGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1410
Dy 602 PheGlySerCysValProHisThrThrArgProLysArgGluTyraGluIleAspGlyArg 621
QY 1411 CCTACTCTATCTTCCACCCAGACGATCATGCGGAAGGATGTCGGCTCCTACGATC 1470
Dy 622 AspTyraHisPheValSerSerArgGluLysMetGluLysAspIleGlnAlaHisLysPhe 641
QY 1471 AAGAAGGAGGATCCTTAGACCTGCGCTGGAAGCGGTGTGGACTCCCCCATTT----- 1524
Dy 642 IleGluAlaGlyGlnTyraAsnSerHisLeuTyraGlyThrSerValGlnSerValArgGlu 661
QY 1525 -----GGAG-----GTGCTGTTTCTGCTGTGTATGAGCGG 1557
Dy 662 ValAlaGluGlnGlyHisCysIleLeuAspValSerAlaAsnAlaValArgArgLeu 681
QY 1558 GGAGCTGCTGAGCGGCTGTCGTCATGTC-----AAAGGAGCAGATC 1602
Dy 682 GlnAlaAlaHisLeuHisProIleAlaIlePheIleArgProArgSerLeuGluAsnVal 701
QY 1603 ATGGAATCAACGCGCAAGATTGTGACAGACTACACCTGGCTGGAGCTGAGCTGCC 1659
Dy 702 LeuGluIleAsnLysArgIleThrGluGlnAlaArgLysAlaPheAspArgAla 720

RESULT 12
US-09-998-425-3
; Sequence 3, Application US/09998425
; Publication No. US20030008346A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavitian, Sean V.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: MMSC1 - An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC1 Gene
; CURRENT APPLICATION NUMBER: US/09/998,425
; CURRENT FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/233,086
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/071,861
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-998-425-3

Alignment Scores:
Pred. No.: 4,99e-11 Length: 1881
Score: 259.50 Matches: 149
Percent Similarity: 37.03% Conservative: 115
Best Local Similarity: 20.90% Mismatches: 232
Query Match: 6.30% Indels: 217
DB: 9 Gaps: 31

```



```
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39113
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005137.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EST_HUMAN HIT: BE266204.1, EVALUOE 5.30e+00
US-09-864-761-39113

Alignment Scores:
Pred. No.: 1.17e-09 Length: 45
Score: 232.00 Matches: 39
Percent Similarity: 88.64% Conservative: 0
Best Local Similarity: 88.64% Mismatches: 5
Query Match: 5.63% Indels: 0
DB: 10 Gaps: 0

US-09-502-945-4 (1-2236) x US-09-864-761-39113 (1-45)
QY 321 TCAGCTGACCCCGCGCGCTCCAGGAGCTGAAGGAGCTGGCTGTGGACGCTGTGCACCC 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 SerProThrSerProValTrpGlnGluAlaGluGlyGlyAlaSerGlyProSerAlaPro 21
QY 381 CGAAGCCTCGCGCTGATGCGTGGTGGCTGGAGTTTGGCTGTGGGCTTTTCATCTC 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22 ArgArgProArgProGluCysAlaTrpTrpProGlyValTrpLeuTipAlaLeuHisLeu 41
QY 441 CCACCTCATCA 452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 42 ProProHisGln 45

Search completed: March 21, 2003, 13:57:52
Job time : 53.9422 secs
```


OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:32:38 ; Search time 39.8301 Seconds
(without alignments)

Title: US-09-502-945-4

Perfect score: 4120
Sequence: 1 cctggcccggtcgcggtgc.....tatttccagcttaaaaaa 2236

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop	10.0	Ygapext	0.5
Ygapop	5.0	Ygapext	7.0

Egapop	6.0	, Egapext	7.0
Delop	6.0	: Delext	7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Command line parameters:

```

Command line parameters:
-MODEL=framet_n2p.model -DEV=xlp
-Q/cgna21/USPTO_spool/US09502945/runat_14032003_101059.19137/app_query.fasta_1.10979
-DB=PIR_73 -FPMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPCPI=0.0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTEXT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=0 -MAXLEN=000000
-USER=US09502945_@CGN_1_1_263_@runat_14032003_101059.19137 -NCPD=6 -ICPU=3
-NO_XLIFY -NO_MMAPP -LARGEXJECT -NEG_SCORE=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6 -FGAPEXT=7
YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : PIR_73:*

```
1: pir1:
```

```
2: pir2:
```

3: pir3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	273.5	6.6	720	2	A45436		synapse-associated
2	269	6.5	2054	2	T46612		multi PDZ domain p
3	263.5	6.4	767	2	T09599		postsynaptic densi
4	261.5	6.3	724	2	JH0800		postsynaptic densi
5	261.5	6.3	1745	2	A48431		tight junction-ass
6	259	6.3	2055	2	T30259		multiple PDZ domai
7	248.5	6.0	926	2	I38756		homolog of prosoph
8	244.5	5.9	904	2	T33757		homolog of Drosoph
9	241.5	5.9	1012	2	T32160		hypothetical prote
10	240.5	5.8	911	2	I56552		synapse-associated
11	230	5.6	870	2	G01974		channel associated
12	227.5	5.5	1131	2	T15617		hypothetical prote
13	225	5.5	852	2	T10811		channel associated
14	221.5	5.4	2450	2	S71625		protein-tyrosine-p

tight junction protein
AMPA glutamate receptor
synaptic scaffolding
tumor necrosis factor
tumor protein
hypothetical protein
atrophin-1 related
eyelid - fruit fly
bazoeka gene protein
protein co-factor
nucleon-binding protein
nucleon-binding protein
tyrosine kinase activator
protein-tyrosine phosphatase
protein-tyrosine phosphatase
brain-specific angiogenesis
gene x104 protein kinase
atypical protein kinase
hypothetical protein
BHLH protein - human
collagen alpha 2(I)
protein-tyrosine phosphatase
probable guanylate kinase
tight junction protein
collagen alpha 1(I)
collagen alpha 2(I)
collagen alpha 2(I)
collagen alpha 1(I)
collagen alpha 1(I)
collagen alpha 1(I)
discs-large tumor suppressor
collagen alpha 4(I)
collagen alpha 4(I)

ALIGNMENTS

RESULT 1

A45436
synapse-associated protein SAP90 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Sep-1993 #sequence_revision 18
C:Accession: A45436
R:Kistner, U.; Wenzel, B.M.; Veh, R.W.; C
J. Biol. Chem. 268, 4580-4583, 1993
A:Title: SAP90, a rat presynaptic protein
A:Reference number: A45436; MUID:93186749
A:Accession: A45436
A:Status: preliminary; not compared with
A:Molecule type: nucleic acid
A:Residues: 1-720 <K15>
A:Experimental source: brain
A:Note: sequence extracted from NCBI back
C:Superfamily: discs-large tumor suppress
F:70-148/Domain: GLGF domain homology <GL
F:165-243/Domain: GLGF domain homology <GL
F:435-493/Domain: SH3 homology <SH3>
F:531-708/Domain: guanylate kinase homolo

Alignment Scores:

Pred. No.:	1.66e-09	Length:	720
Score:	273.50	Matches:	125
Percent Similarity:	40.7%	Conservative:	93
Best Local Similarity:	23.3%	Mismatches:	206
Query Match:	6.64%	Indels:	111
DB:	2	Gaps:	23

US-09-502-945-4 (1-2236) x A45436 (1-720)

QY 340 TCAGGAAGCTGAAGGAGTGCGTCTGGACCGTCTCACCCCAAGCGCTCGGCCGTGAGT 399
::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 155 AlaGLysValMetGLutIleLysLeuIleLys---GlyProLysGlyLeuGLyPheSer 173
::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
QY 400 GTCCGTGGTGGGC-----CTGGAGTTTGGCGTGGGGCTTCATCATCCTCCAC 444
::: ||||||


```

Db 174 lleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 193
QY 445 CTCATCAAGAGCGGTGACGACAGCGCTCGG---CTCCAGGTAGGGGACGACATCGTC 501
Db 194 lleIleGluGlyGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 213
QY 502 CGGATCAATGATATTCATCTCCTCGTACCCATGAGGAGGTCAATCAACCTCATTCGA 561
Db 214 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaLeuLys 233
QY 562 ---ACCAGAAAAGTGTCTCCATCAAGTGTG-----AGACACATCGGCTGATCCCC 609
Db 234 AsnThrTyrAspValValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 253
QY 610 GTGAAAAGCTCTCCTGATGAGCCCTCACTTGG---CAGTATGTGGATCAGTTGTGTGCG 666
Db 254 SerTyrAlaProProAspIleThrThrSerTyrSerGlnHisLeuAspAsnGluIleSer 273
QY 667 GAATCTGGGGGCGTCCGAGCGACCTGGGC----- 696
Db 274 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 289
QY 697 -----TCCCTGGAATCGGAA-----AACAG 720
Db 290 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluAspIleProArg 309
QY 721 GAGAGAAGGCTTTCATCAGCTGCTAGCTCCCGAGCGCTGGCTGAGCATTTCCAGC 780
Db 310 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 329
QY 781 GCCCCATCCAGAACCTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
Db 330 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 348
QY 832 CTGCTGTGCTGAGTGGGATGGAGTGGGACACAGATGTCGAAGTCAATGGCGCTCAGC 891
Db 349 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 366
QY 892 TTCTCTCACTGATCACAAGAGGCTGTAAATGTGCTGAAATAGCGCGACCTGACC 951
Db 367 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyIleThr 385
QY 952 ATCTCCATGTGTAGTCAGCTGCGCGGAGCTGTTTCATCAGACAGCGGAGCGGTGCA 1011
Db 386 ValThrIleIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
QY 1012 GAGCGCGCGCAGCTGAGCTGAGCGGCGAGGAGCTTCTCATCCAGACGGCGTGGCATG 1071
Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
QY 1072 GAGTCCAACAAGATCTCCAGGAGCAGCAGGAG-----ATGGAG 1110
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
QY 1111 CGGCAAGAGGAGAAAGAAATATCCAGAGGCGAGCAGAGGAAATACAGAGATACCGGAG 1170
Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 458
QY 1171 GAGATGGAACACATGTAGAGGAGGAGAGAGT----- 1206
Db 459 ValLeuHisValIleAspAlaGlyAspGluGluTyrTrpGlnAlaArgValHisSer 478
QY 1207 -----AAGAGGAA 1215
Db 479 AspSerGluThrAspAspIleGlyPheIleProSerLysArgValGluArgArgGlu 498
QY 1216 TGGGAAGAAAGACTGGGGCTCA-----AAGAACAGCTACTCTTG 1254
Db 499 TrpSerArgLeuTrpGlySerSerGlySerGlnGlyArgGluAspSerValLeuSer 518
QY 1255 CCTAAACCATCACT---CCTGAGGTACAC-----CCAGTACCCCTTCGCAAGCA 1302
Db 519 TyrGluThrValThrGlnMetGluValHisTyrAlaArgProIleIleLeuGlyPro 538

```

```

QY 1303 AAGTATGATCAGGAGTGGAACTGAGCTGAGCCCGACAGATGATGATGAGGACG 1362
Db 539 ThrLysAspArgAlaAsnAspAspLeuLeuSerProSerProAspLysPheGlySerCys 558
QY 1363 GAGGAGCAGGAGCAGGAGGATTTCGGAATATAGGAAGCTTTGACCCCTACTCTATG 1422
Db 559 ValProHisThrThrArgProLysArgGluTyrGluIleAspGlyArgAspTyrHisPhe 578
QY 1423 TTCACCCAGCAGCAGATCATGCGGAGGATGTCGGGTCTCCATCAGCATCAGAGAGGGA 1482
Db 579 ValSerSerArgGluLysMetGluLysAspIleGlnAlaHisLysPheIleGluAlaGly 598
QY 1483 TCCTTAGACCTGGCCCTGGAAGCGGTGGACTCCCCCAT----- 1524
Db 599 GlnTyrAsnSerHisLeuTyrGlyThrSerValGlnSerValArgGluValAlaGluGln 618
QY 1525 GGAAG-----GTGGTGTGTTCTGCTGTATGAGCGGGAGCTGCTGAG 1569
Db 619 GlyLysHisCysIleLeuAspValSerAlaAsnAlaValArgArgLeuGlnAlaAlaHis 638
QY 1570 CGGCATGGTGCATTGTG-----AAAGGGCAGAGATCATGGCAATCAAC 1614
Db 639 LeuHisProIleAlaIlePheIleArgProArgSerLeuGluAsnValLeuGluIleAsn 658
QY 1615 GGCAGATTGTGACAGACTACACCTGGCTGAGCTGACGCTGCC 1659
Db 659 LysArgIleThrGluGluGlnAlaArgLysAlaPheAspArgAla 673

```

RESULT 2

```

T46612
Multi PDZ domain protein 1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 21-Jul-2000
C:Accession: T46612
R:Ullmer, C.; Schmuck, K.; Figge, A.; Lubbert, H.
FEBS Lett. 424, 63-68, 1998
A:Title: Cloning and characterization of MUPP1, a novel PDZ domain protein.
A:Reference number: 223104; MUID:98196865; PMID:9537516
A:Accession: T46612
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-2054 <ULL>
A:Cross-references: EMBL:AJ001320; NID:g2959978; PIDN:CAA04681.1; PID:g2959979
A:Experimental source: brain
C:Genetics:
A:Gene: MUPP1

```

Alignment Scores:

```

Pred. No.: 4.63e-09 Length: 2054
Score: 269.00 Matches: 120
Percent Similarity: 36.64% Conservative: 72
Best Local Similarity: 22.90% Mismatches: 160
Query Match: 6.53% Indels: 172
DB: 2 Gaps: 17

```

US-09-502-945-4 (1-2236) x T46612 (1-2054)

```

QY 157 GCAGAGAGGAGTATCTCTATGATGCTGGGAATGTACACACAGACCATGGAGCTGGCC 216
Db 1672 AlaThrHisAspGluAlaIleAsnValLeuArgGlnThrProGlnArgVal----- 1688
QY 217 GTGCTCTGGGAGCAGTGAAGCTGTCATCAATGAACCCAGCGCTCTGCTCTGTTGAT 276
Db 1689 -----ArgLeuThrLeuTyr--- 1693
QY 277 GCCATTCCGCGCTGATCCCACTGAAGCACCAGCTGGAATATGATCAGTCAACCCCGG 336
Db 1694 -----ArgAspGluAlaProTyrLysGluAspValCysAspThrPheThr----- 1709
QY 337 CGCTCCAGGAAGCTGAGGAGGCTGCTGAGCGCTGTCACCCCGAGGCGCTCGGCTG 396
Db 1710 -----ValGluLeuGlnLysArgProGlyLysGlyLeuGlyLeu 1722

```



```
QY 397 AGTGTGCGTGGTGGCTGGAGTTGGCTGTGGCTCTTTCATCTCCACCTCATCAAGGC 456
Db 1723 SerIleValGlyLysArgAsn---AspThrGlyValPheValSerAspIleValLysGly 1741
QY 457 GGTGAGGCAGACAGCGCTCGGG---CTCCAGGTAGGGAGGAGATCGTCGGGATCAATGA 513
Db 1742 GlyIleAlaAspAlaAspGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGly 1761
QY 514 TATTTCATCTCTCTGTTACCATGAGGAGGATCATCAACTCATTCGA---ACCAAGAAA 570
Db 1762 GluAspValArgAsnAlaThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGly 1781
QY 571 ACTGTGTCCATCAAACTGAGACACATCGCGCTCATCCCGTGAAGAGCTCTCTGATGAG 630
Db 1782 ThrValThrLeuGluValGlyArgIleLysAlaAlaPheHisSerGluArgArgPro 1801
QY 631 CCCCTCACTGGCAGTATGTGGATCAGTTGTGTGCGAATCTGGGGCGTGGAGGCAGC 690
Db 1802 SerGlnSerSerGlnValSerGluSerSerLeuSerSerPheSerLeuProArgSerGly 1821
QY 691 CTGGGCTCCCTGGAAATCGGAAACAAGGAGAAGAAG-----729
Db 1822 IleHisThrSerGluSerSerGluSerSerAlaLysLysAsnAlaLeuAlaSerGluIle 1841
QY 730 -----GTCTTCATCAGCCTGGTAGGCTCCGAGGCTGGCTGGTGGTGGAGCAT 774
Db 1842 GlnGlyLeuArgThrValGluIleLysLysGlyProAlaAspAlaLeuSerIle 1861
QY 775 TCCAGCGGC-----CCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAACCT 825
Db 1862 AlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetMetHisPro 1881
QY 826 GGCCTCCCTCTCTCAGGTG---GGATTGGAGATAGGGACAGATGTCCAACTCAAT 882
Db 1882 AsnGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleCys 1901
QY 883 GCGTCGACTTCTTAACCTGGATCACAGGAGGCTGTAATGTCTGCTGAANAATAGCGGC 942
Db 1902 GlyThrSerThrAspGlyMetThrIleThrGlnAlaValAsnLeuMetLysAsnAla---1920
QY 943 AGCCTGACCATCTCCATCTAGCTGCAGCTGGCGGGAGCTGTTTCATGACAGCCGGAG 1002
Db 1921 SerGlySerIleGluValGlnValAlaGlyGlyAspVal-----1934
QY 1003 CGGCTGGCAGAGCGCGGAGCGTGAAGTGCAGCGGAGGCTTCTCATGTCAGAGCGG 1062
Db 1935 SerValValThrGlyHisGlnGlnGluLeuAlaAsnProCysLeuAlaPhe-----1951
QY 1063 CTGGCGATGGATCCCAACAGATCTCCAGGAGCAGCAGGATCGAGCGGCAAGGAGA 1122
Db 1952 ThrGlyLeuThrSerSerThrIle-----1959
QY 1123 AAAGAAATGCCAGAGGAGCAGAGGAAATGAGATACCGAAGAGAGATGAACAG 1182
Db 1959 -----1959
QY 1183 ATTGTAGAGGAGAGAGAAGTTTAAGAAGCAATGGGAAGAGACTGGGGCTCAAGGAA 1242
Db 1960 -----PheProAspAspLeuGlyProProGln 1968
QY 1243 CAGCTACTCTTGCCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCA 1302
Db 1969 Ser-----LysThrIleThr-----1973
QY 1303 AAGTATGATCAGGAGTGAACCTGAGCTCGAGCCGCGAGATGACTGGATGGAGGCAG 1362
Db 1974 -----LeuAspArgGlyPro 1978
QY 1363 GAGGACGAGGAGACGAGGATTTCCGAAATATGAGGAAGGCTTGACCCCTCTCTATG 1422
Db 1979 Asp-----1979
```

```
QY 1423 TTCACCCAGACAGCATCATGGGAAGGATGTCGGGCTCTACGCATCAAGAAGAGGGA 1482
Db 1979 -----1979
QY 1483 TCCTTAGACCTGGCCTGGAAGGGGTGTGGACTCCCCCATTTGG-----AAGTGTGTC 1536
Db 1980 GlyLeuGlyPheSerIleValGlyGlyThrGlySerProHisGlyAspLeuProIleTyr 1999
QY 1537 GTTTCGTGTGTATGACGGGGAGCTGCTGAGCGGATGCTGGCATTTGTGAAAGGGAC 1596
Db 2000 ValLysThrValPheAlaLysGlyAlaAlaGluAspGlyArgLeuLysArgGlyAsp 2019
QY 1597 GAGATCATGGCAATCAACGGCAAGATTTGTGACAGACTACACCTGGCTGGCTGACGCT 1656
Db 2020 GlnIleIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaValAla 2039
QY 1657 GCCCTGCAGAAAG 1668
Db 2040 IleLeuLysArg 2043
RESULT 3
T09599
postsynaptic density protein 95 - human
C:Species: Homo sapiens (man)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: T09599
R:Stathakis, D.G.; Hoover, K.H.; You, Z.; Bryant, P.J.
submitted to the EMBL Data Library, July 1998
A:Reference number: Z16761
A:Accession: T09599
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-767 <STA>
A:Cross-references: EMBL:U83192; NID:g3318652; PID:g3318653
A:Experimental source: mammary
C:Genetics:
A:Gene: PSD95
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
F:208-286/Domain: GLGF domain homology <GLG>
F:478-536/Domain: SH3 homology <SH3>
F:578-755/Domain: guanylate kinase homology <GKI>
```

```
Alignment Scores:
Pred. No.: 7.36e-09 Length: 767
Score: 263.50 Matches: 125
Percent Similarity: 40.45% Conservative: 93
Best Local Similarity: 23.19% Mismatches: 206
Query Match: 6.40% Indels: 115
DB: 2 Gaps: 24
```

US-09-502-945-4 (1-2236) x T09599 (1-767)

```
QY 340 TCCAGGAGCTGAAGGAGTGGCTGTGACCGTCTGCACCCCGAGGCGCTGGCTCAGT 399
Db 198 AlaGluLysValMetGluIleLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 216
QY 400 GTGCGTGGTGGC-----CTGGAGTTTGGCTGTGGGCTCTTCATCTCCAC 444
Db 217 IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 236
QY 445 CTCATCAAGCGGTGACGACAGCAGCGTCGGG---CTCAGGTAGGGGAGGAGATCGTC 501
Db 237 IleIleGluGlyGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 256
QY 502 CGGATCAATGATATTCATCTCCTCTGACCCATGAGGAGGTCATCAACCTCATTCGA 561
Db 257 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaLeuLys 276
QY 562 ---ACCAAGAAACTGTGTCTCCATCAAGTG-----AGACACATCGGCCTGATCCCC 609
Db 277 AsnThrThrAspValValThrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 296
QY 610 GTGAAAAGCTCTCTCTGATGAGCCCTCCTCCTTGG---CAGTATGTGGATCAGTTGTGTCG 666
```



```
Db 297 SerTyrAlaProAspIleThrThrSerTyrSerGlnHisLeuAspAsnGluIleSer 316
QY 667 GAATCTGGGGCGCTGCGAGCGAGCTGGGC-----
Db 317 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 332
QY 697 -----TCCCTCGAATCGGAA-----AACAG 720
Db 333 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluGluAspIleProArg 352
QY 721 GAGAAAGGCTTTCATCAGCTGTAGCTCCGAGGCGCTTGGCTGCAGCATTTCCAGC 780
Db 353 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 372
QY 781 GGCCCCATCCAGAAAGCTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
Db 373 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 391
QY 832 CTCTCTGCTGAGTGGAGTGGAGATGAGGGACCATGTTGCAAGTCAATGCGGTGCGAC 891
Db 392 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 409
QY 892 TTCTCTAACTGTCATCACAGGAGGCTGTAATGTCTCTGAAATAAGCCGACCTGACC 951
Db 410 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 428
QY 952 ATCTCCATTGTAGTGCAGCTGGCCGGAGCTGTTTCATGACAGACCGGGAGCGCTGGCA 1011
Db 429 ValThrIleIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 443
QY 1012 GAGCGCGGCGAGCTGAGTGCAGCGCGCAGGAGCTTCTCATCAGAGAAGCGGCTGGCGATG 1071
Db 444 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 461
QY 1072 GAGTCCAAACAAGATCTCCAGGAGCAGCAGGAG-----ATGGAG 1110
Db 462 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 481
QY 1111 CGCAAGAGGAGAAAGAAATTCGCCAAGAGCAGCAGAGGAAATGAGATACCGGAAG 1170
Db 482 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 501
QY 1171 GAGATGGAACAGATTGTAGAGGAGGAGAGAACTTT-----1206
Db 502 ValLeuHisValIleAspAlaSerAspGluGluTrpTrpGlnAlaArgValHisSer 521
QY 1207 -----AAGNAGCAA 1215
Db 522 AspSerGluThrAspAspIleGlyPheIleProSerLysArgArgValGluArgGlu 541
QY 1216 TGGGAA-----GAAGACTGGGGCTCA-----AAGGAA 1242
Db 542 TrpSerArgLeuLysAlaLysAspTrpGlySerSerSerGlySerGlnGlyArgGluAsp 561
QY 1243 CAGCTACTCTTGGCTTAAACCATCACT---GCTGAGGTACAC-----CCAGTACCC 1290
Db 562 SerValLeuSerTyrGluThrValThrGlnMetGluValHisTyrAlaArgProIleIle 581
QY 1291 CTTGCGAAGCCAAAGATGATGATGAGGAGTGGAACTTCAGCTCAGCTCAGCCGAGATGACCTG 1350
Db 582 IleLeuGlyProThrLysArgAlaAsnAspLeuLeuSerGluPheProAspLys 601
QY 1351 GATGGAGGACCGGAGGAGCAGGAGCATTTCCGAAATATGAGGAAGCTTTGAC 1410
Db 602 PheGlySerCysValProHisThrThrArgProLysArgGluTyrGluIleAspGlyArg 621
QY 1411 CCTACTCTATGTTACCCAGACAGATCATCGGGAAGGATCTCCGGTCTCTACGCATC 1470
Db 622 AspTyrHisPheValSerSerArgGluLysMetGluLysAspIleGlnAlaHisLysPhe 641
QY 1471 AAGAGGAGGATCCTTAGACCTGGCCCTGGAAGGGGCTGTGGACATCCCATTT-----1524
```

```
Db 642 IleGluAlaGlyGlnTyrAsnSerHisLeuTyrGlyThrSerValGlnSerValArgGlu 661
QY 1525 -----GGGAAG-----CTGCTCGCTTTCTGCTGTGTATGACGG 1557
Db 662 ValAlaGluGlnGlyLysHisCysIleLeuAspValSerAlaAsnAlaValArgArgLeu 681
QY 1558 GGAGCTGCTAGCGCGCATGTGGCATTTGTG-----AAAGGGACGAGATC 1602
Db 682 GlnAlaAlaHisLeuHisProIleAlaIlePheIleArgProArgSerLeuGluAsnVal 701
QY 1603 ATGGCAATCAACGCAAGATTGTGACAGCATACACCTGGCTGAGCGCTGCGC 1659
Db 702 LeuGluIleAsnLysArgIleThrGluGluGlnAlaArgLysAlaPheAspArgAla 720
RESULT 4
JH0800
postsynaptic density protein PSD-95 - rat
N:Alternate names: brain specific PSD-95 protein; discs-large tumor suppressor protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C:Accession: JH0800; S26407
R:Cho, K.O.; Hunt, C.A.; Kennedy, M.B.
Neuron 9, 929-942, 1992
A:Title: The rat brain postsynaptic density fraction contains a homolog of the drosophila discs-large tumor suppressor protein
A:Reference number: JH0800; MUID:93040233; PMID:1419001
A:Accession: JH0800
A:Molecule type: mRNA
A:Residues: 1-724 <CHO>
A:Cross-references: GB:M96853; NID:g206454; PIDN:AAA41971.1; PID:g206455
A:Experimental source: brain
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
F:70-148/Domain: GLGF domain homology <GLG1>
F:165-243/Domain: GLGF domain homology <GLG2>
F:435-493/Domain: SH3 homology <SH3>
F:535-712/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 9,67e-09 Length: 724
Score: 261.50 Matches: 125
Percent Similarity: 40.45% Conservative: 93
Best Local Similarity: 23.19% Mismatches: 206
Query Match: 6.35% Indels: 115
DB: 2 Gaps: 24
US-09-502-945-4 (1-2236) x JH0800 (1-724)
QY 340 TCCAGAAAGCTGAAGAGGTGCTGCTGCAGCGTCTGCACCCGAGGCGCTGAGT 399
Db 155 AlaGluLysValMetGluIleLysLeuIleLys---GlyProLysGlyLeuGlyPheSer 173
QY 400 GTGCTGTGGTGC-----CTGAGTTTGGCTGTGGGCTCTTCATCTCCAC 444
Db 174 IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 193
QY 445 CTCATCAAGCGGTCAGCAGACAGCTCGG---CTCCAGGTAGGAGCAGATCTGC 501
Db 194 IleIleGluGlyGlyAlaHisLysAspGlyArgLeuGlnIleGlyLysIleLeu 213
QY 502 CGGATCAATGGATATTCATCTCCCTCTACCCATGAGGAGTCAACCTCATTCGA 561
Db 214 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaLeuLys 233
QY 562 ---ACCAAGAAACTGTCTCCATCAAGTG-----AGACATCGGCTGATCCCC 609
Db 234 AsnThrTyrAspValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 253
QY 610 GTGAAAGCTCTCTGATGAGCCCTCACTGG---CAGTATGTCAGTCTTGTGTCG 666
Db 254 SerTyrAlaProProAspIleThrThrSerTyrSerGlnHisLeuAspAsnGluIleSer 273
QY 667 GAATCTGGGGCGTCCGAGGAGCGCTGGC-----696
Db 274 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 289
```



```
Qy 697 -----TCCCTCGAATCGGAA-----AACAG 720
Db 290 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluAspIleProArg 309
Qy 721 GAGAAAGAGTCTTCATCAGCCTGGTAGCTCCCGAGGCTTGCTGCGAGCATTTCCAGC 780
Db 310 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 329
Qy 781 GCCCCATCCAGACCTGGGCATCTTATCACCATTGTG-----AAACCTGGCTCC 831
Db 330 GlyGlu-----AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 348
Qy 832 CTGTCTGTCTGAGTGGGATGGAGATAGGAGGACCATGTCAGATGTCAGTCAATGGCTGCAC 891
Db 349 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 366
Qy 892 TTCTCTAACCTGGATCAACAGGAGCTGTAAATGTGCTGAAATAGCGGAGCCTGACC 951
Db 367 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla-----GlyGlnThr 385
Qy 952 ATCTCCATGTAGCTGAGCTGCGCGGAGCTGTTTCATCAGACCGCGGAGCGCTGGCA 1011
Db 386 ValThrIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
Qy 1012 GAGGCGCGCAGCTGAGCTGCGCGGAGGAGCTTCATGCGAGAACGGCTGCGCATG 1071
Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
Qy 1072 GAGTCCAAACAGATCTCCAGGAGCAGCAGGAG-----ATGCAG 1110
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
Qy 1111 CGGAAAGAGGAAAGAAATGCCCAGAGGAGCAGAGGAGGAGGAGTACCGGAG 1170
Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 458
Qy 1171 GAGATGGAACAGATGTAGAGGAGGAGGAGGAGTTT----- 1206
Db 459 ValLeuHisValIleAspAlaGlyAspGluGluIleTyrPheGlnAlaArgValHisSer 478
Qy 1207 -----AAGAACAA 1215
Db 479 AspSerGluThrAspAspIleGlyPheIleProSerLysArgArgValGluArgGlu 498
Qy 1216 TGGGA-----GAGACTGGGCTCA-----AAGAA 1242
Db 499 TrpSerArgLeuLysAlaLysAspTrpGlySerSerSerGlySerGlnGlyArgGluAsp 518
Qy 1243 CAGTACTCTTCCCTAAACCATCACT---GCTGAGGTACAC-----CCAGTACC 1290
Db 519 SerValLeuSerTyrGluThrValThrGlnMetGluValHisTyrAlaArgProIleIle 538
Qy 1291 CTTCCAGACCAAGATATGATCAGGAGTGGAACCTGAGCTGAGCCCGCAGATGACCTG 1350
Db 539 IleLeuGlyProThrLysAspArgAlaAsnAspAspLeuLeuSerGluPheProAspLys 558
Qy 1351 GATGAGCAGCAGGAGCAGGAGCAGGAGGATTTCCGGAATATGAGGAAGGCTTTGAC 1410
Db 559 PheGlySerCysValProHisThrThrArgProLysArgGluTyrGluIleAspGlyArg 578
Qy 1411 CCCTACTCTATGTTCCACCCAGAGCAGATCATGGGAGGAGGATGCTCGGCTCTCAGCATC 1470
Db 579 AspTyrHisPheValSerSerArgGluLysMetGluLysAspIleGlnAlaHisLysPhe 598
Qy 1471 AAGAGGAGGATCTTAGACTGCGCTGGAAGCGGTGTGGACTCCCGCAT----- 1524
Db 599 IleGluAlaGlyGlnTyrAsnSerHisLeuTyrGlyThrSerValGlnSerValArgGlu 618
Qy 1525 -----GGGAAG-----GTGGCTGTTCTGTGTATGAGCGG 1557
Db 619 ValAlaGluGlnGlyLysHisCysIleLeuAspValSerAlaAsnAlaValArgArgLeu 638
```

```
Qy 1558 GGAGCTCTGAGCGGCATGTGGCATTTGTG-----AAAGGGACGACATC 1602
Db 639 GlnAlaAlaHisLeuHisProIleAlaIlePheIleArgProArgSerLeuGluAsnVal 658
Qy 1603 ATGGCAATCAACGGCAAGATTGTGACAGACTACACCTGGCTGAGGCTGAGCGTCC 1659
Db 659 LeuGluIleAsnLysArgIleThrGluGluGlnAlaArgLysAlaPheAspArgAla 677
RESULT 5
A46431
tight junction-associated protein ZO-1 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A46431
R:Itoch, M.; Nagatuchi, A.; Yonemura, S.; Kitani-Yasuda, T.; Tsukita, S.
J. Cell Biol. 121, 491-502, 1993
A:Title: The 220-kD protein colocalizing with cadherins in non-epithelial cells is id
py.
A:Reference number: A46431; MUID:93252986; PMID:8486731
A:Accession: A46431
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1745 <ITO>
A:Cross-references: GB:D14340; NID:g303709; PIDN:BAA03274.1; PID:d1003784; PID:g30371
A:Experimental source: F9 cells
A:Note: sequence extracted from NCBI backbone (NCBIN:131200, NCBIP:131201)
C:Superfamily: guanylate kinase homology; GLGF domain homology
F:27-106/Domain: GLGF domain homology <GLG1>
F:428-498/Domain: GLGF domain homology <GLG3>
F:645-794/Domain: guanylate kinase homology <GLK>
Alignment Scores:
Pred. No.: 1,31e-08 Length: 1745
Score: 261.50 Matches: 143
Percent Similarity: 36.73% Conservative: 106
Best Local Similarity: 21.09% Mismatches: 222
Query Match: 6.35% Indels: 207
DB: 2 Gaps: 30
US-09-502-945-4 (1-2236) x A46431 (1-1745)
Qy 358 GTGCGTCTGGACCTGCTGCACCCGAGGCTCGCGCTGAGTGTGCTGGCTGGAG 417
Db 24 ValThrLeuHisArgAlaProGlyPheGlyIleAlaIleSerGlyArgAsp 43
Qy 418 -----TTTGCTGTGGG-----CTCTTCATCTCCCACTCAACAGGCGGT 459
Db 44 AsnProHisPheGlnSerGlyGluThrSerIleValIleSerAspValLeuLysGlyGly 63
Qy 460 CAGCAGACAGCGTGGGCTCCAGGTAGGAGCAGATCGTCCGGATCAATGGATATTC 519
Db 64 ProAlaGluGly---GlnLeuGlnGluAsnAspArgValAlaMetValAsnGlyValSer 82
Qy 520 ATCTCTCTCTTACCCATGAGGAGGTCTATCAAC----- 552
Db 83 MetAspAsnValGluHisAlaPheAlaValGlnGlnLeuArgLysSerGlyLysAsnAla 102
Qy 553 -----CTCATTCGAAACCAAGAACTGTGTCCATCAAGTGAGACATCGGCTGATC 606
Db 103 LysIleThrIleArgArgLysLysValGlnIleProValSerHis----- 118
Qy 607 CCCGTGAAAGCTCTCCTGAT---GAGCCCTCACTTGGCAGTATGGGATACATTTGTG 663
Db 119 -----ProAspProGluProValSerAspAsnGluAspAspSerTyrAsp 133
Qy 664 TCGAATCTGGGGCGTCCGAGGAGCAGCTGGCTCCCTGGAATCGGAAACAAAGAG 723
Db 134 GluGluValHisAspProArgAlaGlyArgGlyAlaLeuAlaAsnArgArgSerGlyLys 153
Qy 723 ----- 723
Db 154 SerTrpAlaArgAspArgSerAlaSerArgGluArgSerLeuSerProArgSerAspArg 173
```



```
QY 724 -----AAGAGGTCTTCATCAGCCTGGTAGGCTCCGA 756
Db 174 ArgSerValAlaSerSerGlnProAlaLysProThrLysValThrLeuValLysSerArg 193
QY 757 -----GGCTTGGTGCAGCATTCACGGGCCCATCCAGAACGCTGGC 801
Db 194 LysAsnGluGluTyrglyLeu-----ArgProAla 203
QY 802 -----ATCTTTATCAGCCATCGAAACCTGGCTCCCTGCTGCTGAGGTGGGA---TTG 852
Db 204 SerHisIlePheValLysGluIleSerGlnAspSerLeuAlaAlaArgAspGlyAspIle 223
QY 853 GAGATAGGGGACCATGATGCTGAAGTCAATGCGCTGCGACTCTCTAACTCGATCACAAG 912
Db 224 GlnGluGlyAspValValLeuLysIleAsnGlyThrValThrGluAsnMetSerLeuThr 243
QY 913 GAGGCTGTAAATGCTGCTGAATAATAGCGCAGCCTGACCATCTCCATGTAGTGCAGCT 972
Db 244 AspAlaLysThrLeuIleGluArgSerLysGlyLysLeuLysMetValValGlnArgAsp 263
QY 973 GGCCGGGAGCTGTTTCATG-----990
Db 264 GluArgAlaThrLeuLeuAsnValProAspLeuSerAspSerIleHisSerAlaAsnAla 283
QY 991 ACAGACCGGAGCGGCTGCGACAGCGCGGCAG-----CGTGAG 1029
Db 284 SerGluArgAspAspIleSerGluIleGlnSerLeuAlaSerAspHisSerGlyArgSer 303
QY 1030 CTGACCGCGCAGGACTTCTCATGCAGAGCGGCTG-----1065
Db 304 HisAspArgProArgArgSerGlnSerArgSerProAspGlnArgSerGluProSer 323
QY 1066 -----CGCATGGAGTCCAAAGATCTCCAGGAGCAGCAGGAG 1104
Db 324 AspHisSerThrGlnSerProGlnProSerAsnGlySerLeuArgSerArgGluGlu 343
QY 1105 ATGGAGCGCAAGGAGAAAGAAATTGCCAGAGGCGCAGAGAGAAATAGAGATAC 1164
Db 344 GluArgMetSerLysProGlyAlaIleSerThrProValLysHisValAspAspHisPro 363
QY 1165 CGGAGGAGATGCACAGATT---GTAGAGGAGGAGAGCAAGTTTAAAGACCAATGGGAA 1221
Db 364 ProLysAlaValGluGluValThrValGluLysAsnGluLys-----377
QY 1222 GAAGACTGGGGCTCAAAGGAACAGCTACTCTTGCCTAAACCATCATCTGCTGAGTACAC 1281
Db 378 -----GlnThrPro 380
QY 1282 CCAGTACCCCTTCGAAGCAAGATGATGATCAGGGAGTGGAACTGAGCTCGAG---CCC 1338
Db 381 ThrLeuProGluProLysProValTyraGlnValGlyGlnProAspValAspLeuPro 400
QY 1339 GCAGATGACCTGGAGGCGCAGGAGGAGCAGGAGAGCAGGATTCGGGAATATGAG 1398
Db 401 ValSerProSerAspGlyAlaLeuProAsnSerAlaHisGluAsp-----415
QY 1399 GAAGGCTTTGACCCCTACTCTATGTTTACCCAGACAGATCATGGGAAGGATGTCGGG 1458
Db 416 -----GlyIleLeuArgProSerMetLys 423
QY 1459 CTCCTACGCATCAAGAGGAGGATCCTTTAGACCTGGCCCTGGAAGCGGTGTGGACTCC 1518
Db 424 LeuValLysPheArgLysGlyAspSerValGlyLeuArgLeuAlaGlyLysAsnAsp---442
QY 1519 CCCATTGGGAAGGTGCTTCTGCTGTGTATGACGGGGGAGCTGCTGAGCGGATGCT 1578
Db 443 ---ValGly---IlePheValAlaGlyValLeuGluAspSerProAlaAlaLysGlu---459
QY 1579 GGCATTGTGAAGGGGACGAGATCATGCGCAATCAACGGCAAGATTCTGACAGACTACACC 1638
Db 460 GlyLeuGluGluGlyAspGlnIleLeuArgValAsnAsnValAspPheThrAsnIleIle 479
QY 1639 CTGGCTGAGGCTGACGCTGCCCTGCAGAGGCGCTGGAATCAATGAGCGGGGAGCTGGATCGAC 1698
```

```
Db 480 ArgGluGluAlaValLeuPheLeu-----LeuAsp 489
QY 1699 CTTGTGTTGCGCTCCGCCCAAGGAGTATGACGATGAGCTGACCTTCTTCTCAAG 1758
Db 490 Leu-----ProLys-----GlyGluGluValThrIleLeuAlaGln 501
QY 1759 TCCAAAGGGGAACCAAAATTCACGCTTAGGAACAGTGAAGTCCGCCGCCCTCGT 1818
Db 502 LysLysLys-----AspValTyraArgArgIleVal 511
QY 1819 AACACAAGCCTCGACACGCTTGAG---AGAGGCCACATGACACACACCATGGCAT 1875
Db 512 GluSerAspValGlyAspSerPheTyraIleArgThrHisPheGluTyrgluLysGluSer 531
QY 1876 CTTGGGACCTGGAATCTATCACCCAGGAATCTCAAACTCCCTTTGGCCCTGAACACGAGGC 1935
Db 532 ProTyrgly-----LeuSerPheAsnLysGly 540
QY 1936 CAGATAGGAACAGCTCGGCCACTTTTGAAGGCCAATGTGGAGGAAGGGAGCAGCC 1995
Db 541 GluValPheArgValValAspThrLeuTyraAsnGlyLysLeu---GlySerTrpLeuAla 559
QY 1996 AGCCGTTTGGGAGAGATCTCAAGGATCCAGACTCTCATCTTCCTTCTCT---2043
Db 560 IleArgIleGlyLysAsnHisLysGluValGluArgGlyIleLeuProAsnLysAsnArg 579
QY 2044 -----CTGGCCAGTGAATTTGTCTCTCCCA-----GCTTTGGGGGAC 2082
Db 580 AlaGluGlnLeuAlaSerValGlnTyThrLeuProLysThrAlaGlyGlyAsp 597
RESULT 6
T30259
multiple PDZ domain protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30259
R:Simpson, E.H.; Suffolk, R.; Jackson, I.J.
Genomics 59, 102-104, 1999
A:Title: Identification, sequence, and mapping of mouse multiple PDZ domain protein 9
A:Residues: 1-2055 <SIM>
A:Cross-references: EMBL:AJ131869; NID:g4150877; PIDN:CAA10523.1; PID:g4150878
A:Experimental source: strain C57/BL6 X CBA F1; whole brain
C:Genetics:
A:Gene: mpdz
A:Map position: 4
Alignment Scores:
Pred. No.: 2,01e-08 Length: 2055
Score: 259.00 Matches: 117
Percent Similarity: 35.93% Conservative: 72
Best Local Similarity: 22.24% Mismatches: 161
Query Match: 6.29% Indels: 176
DB: 2 Gaps: 19
US-09-502-945-4 (1-2236) x T30259 (1-2055)
```

```
QY 157 GCAGAGAAGGACTATCTATGATGTGCGAATGTACCACACAGACCATGGAGCTGGCC 216
Db 1673 AlaThrHisAspGluAlaIleAsnValLeuArgGlnThrProGlnArgVal-----1689
QY 217 GTGCTGTGGGAGAGCTGAAGCTGGTCAATGAACCCAGCCGCTGCTGCTCTGTTGAT 276
Db 1690 -----ArgLeuThrLeuTyra---1694
QY 277 GCCATTGCGCGCTGATCCCACTGAGCACCAGCTGGGAATATGATCAGCTGACCCCGG 336
Db 1695 -----ArgAspGluAlaProTyrglyLysGluAspValCysAspThrPheThr-----1710
```


QY 337 CGCTCCAGGAAGCTGAAGGAGGTGGCTGTGGACCGTCTGCACCCCGAAGGCTCGGCGTG 396
 Db 1711 -----lleGluleuGlnLysArgProGlyLysGlyLeuGlyLeu 1723
 QY 397 ACTGTGCGTGGTGGCTGGAGTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGGC 456
 Db 1724 SerIleValGlyLysArgAsn---AspThrGlyValPheValSerAspIleValLysGly 1742
 QY 457 GGTACGAGCAGACGCGTCGGG---CTCCAGGTAGGGGACGAGATCGTCCGGATCAATGGA 513
 Db 1743 GlyIleAlaAspAlaAspGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGly 1762
 QY 514 TATTCCATCTCCTCTGTACCCATCAGGAGGTCATCAACCTCATTCGA---ACCAGGAA 570
 Db 1763 GluAspValArgHisAlaThrGlnGluAlaValAlaLeuLeuLysCysSerLeuGly 1782
 QY 571 ACTGTGTCATCAAAAGTGAGACACATCGCCCTGTATCCCGCTGAAAAAGC----- 618
 Db 1783 AlaValThrLeuGluValGlyArgValLysAlaAlaProPheHisSerGluArgPro 1802
 QY 619 TCTCTGTATGACCCCTCCTTCATGGCAGTATGTGATCAGTTGTTCGGAAATCTGGGGGC 678
 Db 1803 SerGlnSerSerGlnValSerGluSerLeuSerSerPheThrProProLeuSerGly 1822
 QY 679 GTGCGAGGCGCTGGCTCGCTCCCTCGAATCGGAAACAAAGGAGAGAAAGTCTTCATC 738
 Db 1823 IleAsnThrSerGluSerLeuGluSerAsn-----SerLysLysAsnAlaLeuAlaSer 1840
 QY 739 AGCCGTGTAGGCTCCCGA-----GGCCTTGGCTGTC 768
 Db 1841 GluIleGlnGlyLeuArgThrValGluIleLysLysGlyProAlaAspSerLeuGlyLeu 1860
 QY 769 AGCATTTCCAGGGC-----CCCATCCAGAGCCCTGCATCTTTATCAGCATCTGTG 819
 Db 1861 SerIleAlaGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetMet 1880
 QY 820 AAACCTGGCTCCTCTGCTGCTGAGTG---GGATTGGAGATAGGGACACAGATTCGAA 876
 Db 1881 HisProAsnGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThr 1900
 QY 877 GTCATAGCGGTGCTTCTTAACCTGGATCACAAGGAGGTGTAAATGTCTGAAAAAT 936
 Db 1901 IleCysGlyThrSerThrAspGlyMetThrHisThrGlnAlaValAsnLeuMetLysAsn 1920
 QY 937 AGCCGACGCTGACCATCTCCATCTTCTGCTGCTGCGGGGAGCTCTTCTATCAGACAC 996
 Db 1921 Ala---SerGlySerIleGluValGlnValValAlaGlyGlyAspVal----- 1935
 QY 997 CGGAGCGCTGGCAGAGCGCGCGGCGGCTGAGCTGACGCGCAGGAGGCTTCTCATGCAG 1056
 Db 1936 -----SerValValThrGlyHisGlnGlnGluLeuAlaAsnProCysLeuAlaPhe--- 1952
 QY 1057 AAGCGGCTGGCGATGAGTCCCAACAGATCTCCAGGAGCAGCAGGAGATGAGCGGCA 1116
 Db 1953 -----ThrGlyLeuThrSerSerSerIle----- 1960
 QY 1117 AGGAGAAAGAAATGCCCCAGAGCCAGCAGGAGGAAATGAGATATACCCGGAAGAGATG 1176
 Db 1960 ----- 1960
 QY 1177 GAACAGATTGTAGAGGAGGAAGAAGTTTAAAGAGCAATGGGAAGAGCTGGGGCTCA 1236
 Db 1961 -----PheProAspAspLeuGlyPro 1967
 QY 1237 AAGGAACAGCTACTCTTCCCTTAAACCATCACTGCTAGGTATACCCAGTACCCCTTCGC 1296
 Db 1968 SerGlnSer-----LysThrIleThr----- 1974
 QY 1297 AAGCCAAAGTATGATCAGGGAGTGGAACTGAGCTGAGCTGAGCCCGCAGATGACCTGGATGGA 1356
 Db 1975 -----LeuAspArg 1977
 QY 1357 GGCACGGAGGACGGGAGCAGGATTTCCGGAAATATGAGGAAGGCTTGTACCCCTAC 1416

Db 1978 GlyProAsp----- 1980
 QY 1417 TCTATGTTACCCCCAGAGCAGATCATGGGAAGGATGTCGGCTCTACGCATCAAGAAG 1476
 Db 1980 ----- 1980
 QY 1477 GAGGGATCCTTAGACCTGGCCCTGGAAGGGGTGTGGACTCCGCCATTGGG-----AAG 1530
 Db 1981 -----GlyLeuSerPheAsnIleValGlyGlyThrGlySerProHisGlyAspLeuPro 1998
 QY 1531 GTGTCGCTTCTTCTGTGTATGACCGGGGAGCTCTCTGAGCGGATGTCGGCATTTGTGAAA 1590
 Db 1999 IleThrValThrValPheAlaLysGlyAlaAlaGluAspGlyArgLeuLysArg 2018
 QY 1591 GGGCAGCAGATCATGGCAATCAAGCAAGATTTGTGACAGACTACACCTGGCTGGAGGCT 1650
 Db 2019 GlyAspGlnIleIleAlaValAsnGlyClnSerLeuGluGlyValThrHisGluGluAla 2038
 QY 1651 GACGCTGCCCTGCAGAAAG 1668
 Db 2039 ValAlaIleLeuLysArg 2044

RESULT 7

I38756

homolog of Drosophila discs large protein, isoform 2 - human

C:Species: Homo sapiens (man)

C>Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000

C:Accession: I38756

R:Lue, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.

Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994

A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila

A:Reference number: I38756; MUID:95024052; PMID:7937897

A:Accession: I38756

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-926 <RES>

A:Cross-references: EMBL:U013896; NID:g558435; PIDN:AAA50598.1; PID:g558436

C:Genetics:

A:Gene: GDB:DLG1

A:Cross-references: GDB:393278; OMIM:601014

A:Map position: 3q29-3q29

C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h

C:Keywords: alternative splicing; duplication

F:229-307/Domain: GLGF domain homology <GLG1>

F:324-402/Domain: GLGF domain homology <GLG2>

F:588-646/Domain: SH3 homology <SH3>

F:737-914/Domain: guanylate kinase homology <GKI>

Alignment Scores:

Pred. No.:	7.08e-08	Length:	926
Score:	248.50	Matches:	124
Percent Similarity:	38.19%	Conservative:	91
Best Local Similarity:	22.02%	Mismatches:	189
Query Match:	6.03%	Indels:	159
DB:	2	Gaps:	24

US-09-502-945-4 (1-2236) x I38756 (1-926)

QY 340 TCCAGGAAGCTGAAGGAGGTGGCTGTGGACCGTCTGCACCCCGAAGGCTCGGCTGAGT 399

Db 314 SerGlnLysIleMetGluIleLysLeuLys---GlyProLysLeuGlyPheSer 332

QY 400 GTGCGTGGTGGC-----CTGGAGTTTGGCTGTGGGCTCTTCATCTCCAC 444

Db 333 IlealaclyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleThrLys 352

QY 445 CTCATCAAGCGGTGAGGAGCAGCAGCGTGGG---CTCCAGGTAGGGACGAGATCGTC 501

Db 353 IlelecluglyGlyAlaAlaHisLysAspGlyLysGlnleGlyAspLysLeu 372

QY 502 CGGATCAATCGATATCCATCTCTCTCTGATCCCATGAGGAGCTCATCAACCTCATCGA 561

Db 373 AlaValAsnAsnValCysLeuGluGluValThrHisGluGluAlaValThrAlaLeuLys 392
 QY 562 ---ACCAAGAAACTGCTGCCATCAAGTG-----AGACACATC 597
 Db 393 AsnThrSerAspPheValTyrLeuLysValAlaLysProThrSerMetTyrMetAsnAsp 412
 QY 598 GCGCTGATCC---GTGAAAAGCTCTCTGTATGAGCCCTCCTCATTGGCAGTATGTG 651
 Db 413 GlyTyrAlaProAspIleThrAsnSerSerGlnPro-----Val 427
 QY 652 GATCAGTTTGTCTCGGAATCTGGGGCGTCGAGCAGCCTGGCTCCCTC----- 702
 Db 428 AspAsnHisValSerProSerPheLeuGlyGlnThrProAlaSerProAlaArgTyr 447
 QY 703 -----GGAATCGGGAA---AACAGGAGAGAAAGTCTTC 735
 Db 448 SerProValSerLysAlaValLeuGlyAspAspGluIleThrArgGluProArgLysVal 467
 QY 736 ATCAGCCTGGTAGCTCCGAGGCTTGGCTGCAGATTTCAGCGGCCCATCCAGAG 795
 Db 468 ValLeuHisArgGlySerThrGlyLeuGlyPheAsnIleValGlyGlu---AspGly 486
 QY 796 CCTGGCATCTTATCAGCCATGTG-----AACTGGCTCCCTCTCTCTCAGGTG 846
 Db 487 GluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu--- 505
 QY 847 GGATTGGAGATAGGACCAAGTTCGAAAGTCAATGGCTCGACTCTCTAACCTGGAT 906
 Db 506 ---LeuArgLysGlyAspArgIleIleSerValAsnSerValAspLeuArgAlaAlaSer 524
 QY 907 CACAGGAGCTGTAATGTCTGAAATAAGC---CGCAGCTGACCATC----- 954
 Db 525 HisGluGlnAlaAlaAlaLeuLysAsnAlaGlyGlnAlaValThrIleValAlaGln 544
 QY 954 ----- 954
 Db 545 TyrArgProGluGluTyrSerArgPheGluAlaLysIleHisAspLeuArgGluGlnMet 564
 QY 955 -----TCCATTGTAGCTGCAGCTGGC----- 975
 Db 565 MetAsnSerSerIleSerSerGlySerGlySerLeuArgThrSerGlnLysArgSerLeu 584
 QY 976 -----CGGAGCTGTTCATGACAGACCGGGCCGCTGGCA----- 1011
 Db 585 TyrValArgAlaLeuPheAspTyrAspLysThrLysAspSerGlyLeuProSerGlnGly 604
 QY 1011 ----- 1011
 Db 605 LeuAsnPheLysPheGlyAspIleLeuHisValIleAsnAlaSerAspGluTrpTrp 624
 QY 1012 GAGCGCGGCAG----- 1023
 Db 625 GlnAlaArgGlnValThrProAspGlyGluSerAspGluValGlyValIleProSerLys 644
 QY 1024 CTGAGCTGCAGCGGAGGCTTCTATGCAAGACGGCTGGATGGAGTCCCAACAG 1083
 Db 645 ArgArgValGluLysLysGluArgAlaArgLeuLysPheAsnSer---Lys 663
 QY 1084 ATCCTCAGGAGCAGCAGAGATGGAGCGCAAGAGAGAAAGAAATGCCCCAG----- 1137
 Db 664 ThrArgAspLysGlyGlnSerPheAsnAspLysArgLysLysAsnLeuPheSerArgLys 683
 QY 1138 -----AAGGCAGCAGGAGAAATGAGATACCGGAGAGGAGATGGAAACAGATT 1185
 Db 684 PheProPheTyrLysAsnLysAspGlnSerGluGlnGluThrSerAspAlaAspGlnHis 703
 QY 1186 GTA-----GAGGAGGAGAGAAAGTTTAAGAACGAATGGGAGAGACTGG 1230
 Db 704 ValThrSerAsnAlaSerAspSerGluSerSerTyrArg----- 716
 QY 1231 GGCTCAAGCAAGCTACTCTTGGCTTAAACCATCACTGCTGAG-----GTA 1278
 Db 717 GlyGlnGluGluTyrValLeuSerTyrGluProValAsnGlnGlnGluValAsnTyrThr 736

QY 1279 CACCAGTACCCCTTCGCAAGCAAAAGTATGATCAGGAGTGGAACCTGAGCTCGAGCCC 1338
 Db 737 ArgProValIleIleLeuGlyProMetLysAspArgIleAsnAspLeuIleSerGlu 756
 QY 1339 GCAGATCACCTGGATGGAGGACCGGAGGAGCAGGAGGATTTCCGGAATATGAG 1398
 Db 757 PheProAspLysPheGlySerCysValProHisThrThrArgProLysArgAspTyrGlu 776
 QY 1399 GAAGGCTTTGACCCCTACTCTATGTTTCCACCCAGAGAGATCATCGGGAAGATGTCGG 1458
 Db 777 ValAspGlyArgAspTyrHisPheValThrSerArgGluGlnMetGluLysAspIleGln 796
 QY 1459 CTCCTACCATCAAGAGGAGGATCCTTAGACCTGCCCTGGAGGCGGTGGACTCC 1518
 Db 797 GluHisLysPheIleGluAlaGlyGlnTyrAsnAsnHisLeuTyrGly----- 812
 QY 1519 CCCATTGGGAAGTGGCTGTTCTGCTGTATGAGCGGGAGCTGTGAGCGGCATGGT 1578
 Db 813 -----ThrSerValGlnSerValArgGluValAlaGlyLysGlyLysHisCys 828
 QY 1579 GGCATTGTGAAAGGAGCAGATCATGCAATCAACGGCAAGATTGTGACAGACTACACC 1638
 Db 829 -----IleLeuAspValSerGlyAsnAlaIleLysArgLeuGln 841
 QY 1639 CTGGCTGAG 1647
 Db 842 IleAlaGln 844
 RESULT 8
 I38757
 homolog of Drosophila discs large protein, isoform 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
 C:Accession: I38757
 R:Luë, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.
 Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
 A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila
 A:Reference number: I38756; MUID:95024052; PMID:7937897
 A:Accession: I38757
 A:Molecule type: mRNA
 A:Status: preliminary
 A:Residues: 1-904 <RES>
 A:Cross-references: EMBL:U13897; NID:g558437; PIDN:AAA50599.1; PID:g558438
 C:Genetics:
 A:Gene: GDB:DLG1
 A:Cross-references: GDB:393278; OMIM:601014
 A:Map position: 3q29-3q29
 C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
 C:Keywords: alternative splicing; duplication
 F:229-307/Domain: GLGF domain homology <GLG1>
 F:324-402/Domain: GLGF domain homology <GLG2>
 F:588-646/Domain: SH3 homology <SH3>
 F:715-892/Domain: guanylate kinase homology <GKI>
 Alignment Scores:
 Pred. No.: 1.26e-07 Length: 904
 Score: 244.50 Matches: 121
 Percent Similarity: 38.88% Conservative: 87
 Best Local Similarity: 22.62% Mismatches: 202
 Query Match: 5.93% Indels: 125
 DB: 2 Gaps: 22
 US-09-502-945-4 (1-2236) x I38757 (1-904)
 QY 340 TCCAGGAGCTGAAGGAGTGGCTGCTGACCGCTCTGCACCCGAGGCTCGGCCTGAGT 399
 Db 314 SerGluLysIleMetGluIleLysLeuLys---GlyProLysGlyLeuGlyPheSer 332
 QY 400 GTGCGTGGTGGC-----CTGGCTGTGGCTCTTCATCTCCAC 444
 Db 333 IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 352

QY	445	CTCATCAAGAGCGCTCAGGCAGACAGCGTGGG---CTCAGAGTAGGGGACAGATCGTC	501
Db	353	IleIleGluGlyAlaAlaHisLysAspGlyLysLeuGlnIleGlyAspLysLeu	372
QY	502	CGGATCAATGGATTCATCTCCTCTGTACCCATGAGGAGTCAACCTCATTCGA	561
Db	373	AlaValAsnAsnValCysLeuGluValThrHisGluGluAlaValThrAlaLeuLys	392
QY	562	---ACCAAGAAACGTGGTCCATCAAGTG-----AGACATCATC	597
Db	393	AsnThrSerAspPheValTyrLeuLysValAlaLysProThrSerMetTyrMetAsnAsp	412
QY	598	GGCTGATCCCC-----GTGAAAGCTCTCTCTGTAGAGCCCTCACTTGGCAGTATGTG	651
Db	413	GlyTyrAlaProProAspIleThrAsnSerSerSerGlnPro-----Val	427
QY	652	GATCAGTTTGTGTCGGAATCTGGGGCGTGCAGGCAGCTGGGTCCCTC-----	702
Db	428	AspAsnHisValSerProSerSerPheLeuGlyGlnThrProAlaSerProAlaArgTyr	447
QY	703	-----GGAAATCGGGA---AACAAAGGAGAAGAAGCTCTTC	735
Db	448	SerProValSerLysAlaValLeuGlyAspAspGluIleThrArgGluProArgLysVal	467
QY	736	ATCAGCTGTGTAGCTCCCGAGCGCTTGCTGCAGCATTTCCAGCGGCCCATCCAGAA	795
Db	468	ValLeuHisArgGlySerThrGlyLeuGlyPheAsnIleValGlyGlu---AspGly	486
QY	796	CTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCCTCTCTCCTCAGGTG	846
Db	487	GluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu---	505
QY	847	GGATTGGAGATAGGGGACCAAGATGTGCAAGTCAATGGCTGCAGCTTCTAACTGGAT	906
Db	506	---LeuArgLysGlyAspArgIleIleSerValAsnSerValAspLeuArgAlaLys	524
QY	907	CACAAGAGGCTGTAAATGTCTGAAAATAGCCGCGAGCTGCACCATCTCCATTGTAGCT	966
Db	525	HisGluGlnAlaAlaAlaAlaLeuLysAsnAla---GlyGlnAlaValThrIleValAla	543
QY	967	GCAGTGGCCGGAGCTGTTTCATGCAGACCGCGGAGCGGCTGGCAGCGCGCCAG---	1023
Db	544	GlnTyrArgProGluGluTyrSerArgPheGluAlaLysIleHisAspLeuArgGluGln	563
QY	1024	-----CGTGAGCTCGACGCGGCAGGAG	1044
Db	564	MetMetAsnSerSerIleSerSerGlySerGlySerLeuArgThrSerGlnLysArgSer	583
QY	1045	CTTCTCATGCAGAGCGGCTGGCGATGGAGTCCACAAG-----	1083
Db	584	LeuTyrValArgAlaLeuPheAspTyrAspLysThrLysAspSerGlyLeuProSerGln	603
QY	1084	-----ATCCCTC-----	1089
Db	604	GlyLeuAsnPheLysPheGlyAlaPheIleLeuHisValIleAsnAlaSerAspGluTyr	623
QY	1090	---CAGGACGACGAGAGATGGAGCGGCAAGAGGAGAAAGAAATAT-----GCC	1134
Db	624	TrpGlnAlaArgGlnValThrProAspGlyGluSerAspGluValGlyValIleProSer	643
QY	1135	CAGAAGCGACGAGGAAATGAGAGATACCGGAGGAGAGATGGAACAGATTGTAGAGGAG	1194
Db	644	LysArgArgValGluLysLysGluArgAlaArgLeuLysThrValLysPhe---AsnSer	662
QY	1195	GAAGAGAAGTTTAAAGACCAATGGGAAGAGACTGGGCTCAAG-----	1239
Db	663	LysThrArgAspLysGlyGluIleProAspAspMetGlySerLysGlyLeuLysHisVal	682
QY	1240	-----GAACAGCTACTCTTG	1254
Db	683	ThrSerAsnAlaSerAspSerGluSerSerTyrArgGlyGlnGluGluTyrValLeuSer	702
QY	1255	CCTAAACCATCATCTGCTGAG-----GTACACCCAGCTACCCCTTCGCAAGCCA	1302

Db 703 TyrGluProValAsnGlnGlnValAsnTyrThrArgProValIleIleLeuGlyPro 722
 Qy 1303 AAGTATGATCAGGAGTGGAAACCTGAGCTCGAGCCCGCATGATGCTGATGAGGACG 1362
 Db 723 MetLysAspArgIleAsnAspLeuIleSerGluPheProAspLysPheGlySerCys 742
 Qy 1363 GAGGACGAGGAGCAGCAGGATTCGCGAAATATAGGAAAGCGTTTGACCCCTACTCTATG 1422
 Db 743 ValProHisThrThrArgProLysArgAspTyrGluValAspGlyArgAspTyrHisPhe 762
 Qy 1423 TTCACCCAGCAGCAGATCATGCGGAGGATGTCGCGCTCTACCATCATCAAGAGGAGGA 1482
 Db 763 ValThrSerArgGluGlnMetGluLysAspIleGlnGluHisLysPheIleGluAlaGly 782
 Qy 1483 TCCTTAGACCTGGCCCTCGGAAGCGGTGTGGACTCCCCCATTTGGGAAGGTGCTGTTTCT 1542
 Db 783 GlnTyrAsnAsnHisLeuTyrGly-----ThrSerValGln 794
 Qy 1543 GCTGTGTATGAGCGGGAGCTGCTGAGCGGCATGTGGCATTTGTGAAGGGGACGAGATC 1602
 Db 795 SerValArgGluValAlaGlyLysGlyLysHisCys-----Ile 807
 Qy 1603 ATGGCAATCAACGCGAAGATTGTGACAGACTACACCTGGCTGAG 1647
 Db 808 LeuAspValSerGlyAsnAlaIleLysArgLeuGlnIleAlaGln 822
 RESULT 9
 T23160
 hypothetical protein K01A6.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
 C:Accession: T23160
 R:Cottage, A.
 submitted to the EMBL Data Library, January 1996
 A:Reference number: Z19701
 A:Accession: T23160
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1012 <SWIL>
 A:Cross-references: EMBL:Z68750; PIDN:CAA92965.1; GSPDB:GN00022; CESP:K01A6
 A:Experimental source: clone K01A6
 C:Genetics:
 A:Gene: CESP:K01A6.1
 A:Map position: 4
 A:Introns: 107/1; 151/2; 183/2; 270/2; 316/3; 410/3; 628/3; 922/3; 971/3
 C:Superfamily: WW repeat homology
 F:131-168/Domain: WW repeat homology <WWR>
 Alignment Scores:
 Pred. No.: 2,04e-07 Length: 1012
 Score: 241.50 Matches: 121
 Percent Similarity: 34.84% Conservative: 72
 Best Local Similarity: 21.84% Mismatches: 194
 Query Match: 5.86% Indels: 167
 DB: 2 Gaps: 19
 US-09-502-945-4 (1-2236) x T23160 (1-1012)
 Qy 29 CCAGCTCCTGGCAGCGGCGACCGAAGACGGGTCTG-----GCAAGCAGC 76
 Db 574 ProlysThrArgSerArgThrProSerAlaAlaPheArgTyrGlyGluProGlnThrAsn 593
 Qy 77 CAGCTGGACCTGGCCCGACCA-----TGAC 103
 Db 594 MetMetAspSerAlaAlaProLeuProValArgSerLysThrProAlaGluArgGlnThr 613
 Qy 104 GAAAAGTGGCCGAGAAATCCGGCATAAGGTGGATTTTCTGATTGAAATGATGCAGAGA 163
 Db 614 SerArgThrGluGluAspGlnAsnValArg-----AsnThrLeuGlnArg 628
 Qy 164 AGGACATATCTCTATGATGTGCTGCGAAATGTACCAACGACCATGGCGTGGCTGCG 223


```
Db 629 Gln-----ProAlaValThrSerGluTrp-GluGlyMetSerSe 641
Qy 224 TGGGAGACCTGAAGTGTCTCATCAATGAACCCAGCGCTGCTCTGTTTATGTCATTC 283
Db 641 AlaIlePro---AlaSerArgMet-ArgProSerSerThrThrLeuGlyPheAlaThrP 660
Qy 284 GCCCCTGATCCACTGAAG-----CACAGGTGGAATATGATCAGTGCACCCCGGC 337
Db 660 roAsnTyIleProLeuSerGlnTyAsnGlnLysProSerAspLeuIleThrValSerL 680
Qy 338 GCTCCAGGAAGCTGAAGGAGTGTCTGGACCGTCTGCACCCGGAAGCGCTCGCCTGA 397
Db 680 euIleArgLys-----ProValGlyPheGlyPheA 690
Qy 398 GTGCGGTGGTGGCTGAGTTGGCTGTTCATCTCCACCTCATCAAGGCG 457
Db 690 rgLeuLeuGlyValGluSerLysThrProLeuSerValGlyGlnIleValIleGlyG 710
Qy 458 GTCAGGCAGACAGCTCGG-----CTCCAGGTAGGGAGAGATCTCCGGATCAATGAT 514
Db 710 lyAlaAlaGluAspGlyArgLeuGlnGluGlyAspGluIleValGluIleAspGlyH 730
Qy 515 ATTCACTCTCCCTGTACCCATGAGGAGTCAATCAACCTCATTCGAACCAAGAAACTG 574
Db 730 isAsnValGluClyAlaSerHisSerGluAlaValLeuLeuGlu-----A 746
Qy 575 TGTCCATCAAGTGAACACATCGCCCTGATCCCGTGAAGAGCTCTCTGATGAGCC 634
Db 746 laAlaAlaGlnAsnLysHisValLysLeuIleValArgArgProSerArgThrAspProA 766
Qy 635 TCACITGGCAGTATGTGGATCAGTTGTGCGAATCTGGGGCGTGGCGGACCCCTGG 694
Db 766 laArg-----ArgGlySerLeuA 772
Qy 695 GCTCCCTCGAATCGGAAACAAAGGAGAGAGTCTTCATCAGCTGTGAGCTGCC 754
Db 772 snSerAlaGlyProSerGlySerTyAsp-----ValLeuLeuHisArgAsnGluAsnA 790
Qy 755 GAGGCTTGGTGCAGCATTTCCAGCGGCCCATCCAGAAAGCTTGCATCTTTATCAGCC 814
Db 790 spGlyPheGlyPheValLeuMetSerSerGln---HisLysAsnGlySerThrValGlyG 809
Qy 815 ANGTGAACCTCCCTGCTGCTGAGGTGGA---TTGCAGATAGGGAGACCATG 871
Db 809 InIleGlnProGlySerProAlaAlaArgCysGlyArgLeuSerValGlyAspArgValI 829
Qy 872 TCGAAGTCAATGGCTCGACTTCTTAACCTGGATCAAGAGCGCTGAATGTCTGA 931
Db 829 leAlaValAsnGlyIleAspIleLeuSerLeuSerHisProAspThrIleSerLeuIleL 849
Qy 932 AAATAGCCGC-----AGCCTGACCATCTCCATTTGATGCTGCGCGCGGAGC 982
Db 849 ysAspSerGlyLeuSerValArgLeuThrIleAlaProAsnThrAlaGlyProValI 869
Qy 983 TGTTCATGACACCGGGAGCGGTGGCAGAGCGCGGAGCTGAGCTGCAGCGCGCAGG 1042
Db 869 euProMetVal-----S 873
Qy 1043 AGCTTCTCATGACAGAGCGGTGGCGATGGAGTCCAACAAGATCTCTCCAGGAGCAGCAGG 1102
Db 873 erAlaThrLeuGlyArgAsnPheThrMetAsnGlyHis----- 885
Qy 1103 AGATGAGCGGCAAGGAGAAAGAAATTTGCCAGAGGACGACAGAGGAAATGAGAT 1162
Db 885 ----- 885
Qy 1163 ACCGGAAGGAGATGGACAGATTGTAGAGGAGAGAGAACTTTAAGANGCAATCGGAG 1222
Db 886 -----TyGIus 888
Qy 1223 AAGACTGGGGCTCAAGGAACAGCTACTCTTGCCTAAACCATCACTGCTGAGGTACACC 1282
Db 888 erAsnTyrgly-----LeuProProProSerValTyGIusHisP 904
```

```
Qy 1283 CAGTACCCCTCCGCAAGCCAAAGTATGATCAGGAGTGAACCTGAGCTCGAGCCCGCAG 1342
Db 904 roProProSerTyIleAlaPheAspGlyLeuSerIleAsnAspArgMetSerMetAsnG 924
Qy 1343 ATGACCTGGATGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1402
Db 924 lyAsnLeuIleAspValThrLeuGluArgGlyThrLysGlyPhe-----G 939
Qy 1403 GCTTTGACCCCTACTCTATGTTTCCACCCAGAGCAGATCATGGGGAAGGATGTCGCGCTCC 1462
Db 939 lyPhe----- 940
Qy 1463 TACGCATCAAGAGGAGGATCTTAGACCTGGCCCTGGAGCGGTTGGACTCCCGCCA 1522
Db 941 -----SerIleArgGlyGlnGluPheGlyS 950
Qy 1523 TTGGGAAGTGGTCTGTTCTGCTGTATGAGCGGGAGCTGTGAGCGGCATCGTGGCA 1582
Db 950 erMetProLeuPheValLeuArgIleAlaAspGlyProAlaLysAlaAspGlyArgL 970
Qy 1583 TTGTGAAGGAGGAGCAGATCATGCAATCAACGGAAG 1620
Db 970 euGlnValGlyAspGlnLeuThrThrIleAsnGlyGln 982
RESULT 10
I56552
synapse-associated protein 97 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jan-2000
C:Accession: I56552
J:Mueller, B.M.; Kistner, U.; Veh, R.W.; Cases-Langhoff, C.; Becker, B.; Gundelfinger,
R. Neurosci. 15, 2354-2366, 1995
A:Title: Molecular characterization and spatial distribution of SAP97, a novel presyn
A:Reference number: I56552; MUID:95198112; PMID:7891172
A:Accession: I56552
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-911 <RES>
A:Cross-references: EMBL:U14950; NID:9642455; PIDN:AAA79976.1; PID:9642456
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
F:229-307/Domain: GLGF domain homology <GLGI>
F:323-401/Domain: GLGF domain homology <GLG2>
F:587-645/Domain: SH3 homology <SH3>
F:722-899/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 2,27e-07 Length: 911
Score: 240.50 Matches: 119
Percent Similarity: 38.31% Conservative: 89
Best Local Similarity: 21.92% Mismatches: 200
Query Match: 5.84% Indels: 135
DB: 2 Gaps: 22
US-09-502-945-4 (1-2236) x I56552 (1-911)
Qy 343 AGGAAGCTGAAGGAGTGTGCTGACCGCTGACCCCGAAGCGCTGCGCTGAGTGTG 402
Db 314 ArgLysAsnHisGluIleLysLeuIleLys---GlyProLysGlyLeuGlyPheSerIle 332
Qy 403 CGTGGTGGC-----CTGGAGTTTGGCTGTGGCTCTTCATCTCCACCTC 447
Db 333 AlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyIleValThrLysIle 352
Qy 448 ATCAAGCGCGTGAAGGAGCAGACGCTCGGG---CTCCAGGTAGGAGCAGATCGCTCGG 504
Db 353 IleGluGlyGlyAlaAlaHisLysAspGlyLysLeuGlnIleGlyAspLysLeuLeuAla 372
Qy 505 ATCAATGATATTCATCTCTCTCTGTACCATGAGGAGGTGTCATCAACCTCATTCGA--- 561
Db 373 ValAsnSerValCysLeuGluGluValThrHisGluAlaValThrAlaLeuLysAsn 392
Qy 562 ACCAAGAAACTGTGTCATCAAGTGAGA-----CACATCGGC 600
```


Db	743	ProAspLysPheGlySerCysValProHisThrThrArgProLysArgAspTyrGluVal	763
Qy	1402	GGCTTTGACCCCTACTCTATGTTTCCACCCAGAGCAGATCATGGGAAGCATGTCGGGCTC	1461
Db	763	AspGlyArgAspTyrHisPheValThrSerArgGluGlnMetGluLysAspIleGlnGlu	782
Qy	1462	CTAGGCATCAAGAAGGAGGATCTTTAGACCTGGCCCTGGGAAGGC---GGTGTGGACTCC	1518
Db	783	HisLysPheIleGluAlaGlyGlnTyrAsnAsnHisLeuTyrGlyThrSerValGlnSer	802
Qy	1519	CCCATGGGAAGTGGTCTGCTGCTGTGTATGACGGGGAGCTGCTGACGGGCATGGT	1578
Db	803	-----ValArgAlaValAlaGluLysGly-----LysHisCys	813
Qy	1579	GGCATTGTGAAGGGGACGAGATCATGGCAATCAACGCCAAGATTGTGCACAGACTACACC	1638
Db	814	-----IleLeuAspValSerGlyAsnAlaIleLysArgLeuGln	826
Qy	1639	CTGGCTGAG	1647
Db	827	IleAlaGln	829
RESULT 11			
G01974			
channel associated protein of synapse - human			
C:Species: Homo sapiens (man)			
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 21-Jan-2000			
C:Accession: G01974			
R:Kim, E.; Cho, K.			
submitted to the EMBL Data Library, July 1995			
A:Reference number: G08966			
A:Accession: G01974			
A:Status: preliminary; translated from GB/EMBL/DDBJ			
A:Molecule type: mRNA			
A:Residues: 1-870 <KIM>			
A:Cross-references: EMBL:U32376; NID:g1463025; PTD:g1036790			
C:Genetics:			
A:Gene: chapsyn-110			
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylat			
F:198-276/Domain: GLGF domain homology <GLG2>			
F:543-601/Domain: SH3 homology <SH3>			
F:681-858/Domain: guanylate kinase homology <GKI>			
Alignment Scores:			
Pred. No.:		1.04e-06	Length: 870
Score:		230.00	Matches: 143
Percent Similarity:		35.74%	Conservative: 114
Best Local Similarity:		19.89%	Mismatches: 222
Query Match:		5.58%	Indels: 240
DB:		2	Gaps: 32
US-09-502-945-4 (1-2236) x G01974 (1-870)			
Qy	208	GAGCTGGCGCTGCTCTGGGAGACGTAGCTGTCATCAATGAA-----CCACGCGGT	261
Db	141	AspGlyArgLeuArgValAsnAspCysIleLeuArgValAsnGluValAspValSerGlu	160
Qy	262	CTGCCTCTGTTGATGCGCAATTCGGCGGCTGATCCCATCTGAAGCACCAGGTGAATATGAT	321
Db	161	ValSerHisSerLysAlaValGlu-----AlaLeuLysGluAlaGlySerIleAla	177
Qy	322	CAGCTGACCCCGCGGCTCCAGGAAGCTG-----AAGGAGGTGCGTCTGGAC	369
Db	178	ArgLeuTyrValArgArgArgProIleLeuGluThrValValGluIleLysLeuPhe	197
Qy	370	CGTCTGACCCCGAAGCTCTGGCTGAGTGTGCTGTGGC-----CTG	414
Db	198	Lys----GlyProLysGlyLeuGlyPheSerIleAlaGlyGlyValGlyAsnGlnHisIle	216
Qy	415	GAGTTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGCGGTGACGACAGCGCTC	474
Db	217	ProGlyAspAsnSerIleTyrValThrLysIleAspGlyGlyAlaAlaGlnLysAsp	236

Percent Similarity:	34.30%	Conservative:	86
Best Local Similarity:	20.39%	Mismatches:	20
Query Match:	5.52%	Indels:	19
DB:	2	Gaps:	22

US-09-502-945-4 (1-2236) x T15617 (1-1131)

151	QY	AATGATCGACGAGAGGACTATCTCTATGATGTGCTGCGAATGTACACACGACACCATGGAC	210	Db	AsnSerSerThrHisTyrLeuHisGluArg	278	QY	GTGGCCGTGCTCGTGGGAGACCTGAAGCTGGTCATCAATGAACCCAGCCGCTGCGCTCTG	270	Db	AspGlyThrTrpArgGluThrThrArgThrValAspThrProserGlyLeuGlu---	278	QY	TTTGATGCCATTCGGCCGCTGATCCCACTGAAGCACACAGGTGGAAATATGATCAGCTGACC	330	Db	279	-----ArgArgValValGluHisThrGlyValIle	288	QY	CCCCGGGCTCCAGGAG-----CTGAAGAGGTGGCTGTGACCGTGTGCACCCCGAA	384	Db	289	AspAspHisGlyArgLysTrpGluLeuGluAsnIleValLeuGluLysGlyHis---Thr	307	QY	385	GGCTCGGCTGAGTGTGCGTGGTGGCGTG-----GAGTTTGGCTGTGGG	429	Db	308	GlyLeuGlyPheSerIleThrGlyGlyMetAspGlnProThrGluAspGlyAspThrSer	327	QY	430	CTCTTCATCTCCCACTCATCAAGGGGTGACGACAGACGTCGGG---CTCCAGGTA	486	Db	328	IleTyrValThrAsnIleLeuGluGlyGlyAlaAlaLeuAlaAspGlyArgMetArgLys	347	QY	487	GGGACGAGATGCGCGGATCAATGGATATCCCATCTCTCTCTGATCCCATGAGGAGGTC	546	Db	348	AsnAspIleIleThrAlaValAsnAsnThrAsnCysGluAsnValLysHisGluValAla	367	QY	547	ATCAACCTCATTCGAACCAAGAAACTGTG-----TCCATCAAGTGTAGA-----	591	Db	368	ValAsnAlaLeuLysSerSerGlyAsnValSerLeuLysArgArgLysAsp	387	QY	591	-----	591	Db	388	GluAlaPheLeuProIleGlyGlyAsnPheGlyGlySerThrSerTyrLeuArgSerGly	407	QY	592	-----CACATCGCCCTGATCCCC	609	Db	408	ValThrProSerValSerAlaGlyAsnLeuGluHisAlaIleHisSerProSerAlaPro	427	QY	610	GTGAAAAGCTCTCCTGTATGAGCCCTCCTGATGGCATGTATGTGATCAGTATTGTGTCGAA	669	Db	428	IleHisProProProProProValHisHisGlySerLeuSerGlnLeu-----	444	QY	670	TCGGGGCGTGGAGCACCTGGCTCCCTGGAAATCGGAAACAAAGGAGGAAG	729	Db	445	-----SerValGlyGln-----TyrArgSerThrArgProAsnThr	456	QY	730	GTCTTCATCAGCCTGTA---GGCTCCGAGGCTTGGTGTGACAGTAATTCACGCGCC	786	Db	457	SerValIleAspLeuValLysGlyAlaArgGlyLeuGlyPheSerIleAlaGlyGln	476	QY	787	ATCCAGAAG-----CCTGGCATCTTTATCAGCCAT---GTGAAACCTGCG	828	Db	477	GlyAsnGluHisValLysGlyAspThrAspIleTyrValThrLysIleIleGluGluGly	496	QY	829	TCCCTGTCTGCTGAGGTGGATTCGAGATAGGGGACCATGATCTCAAGATCAATGCCGTC	888	Db	497	AlaAlaGluLeuAspGlyArgLeuArgValGlyAspLysIleLeuGluValAspHisHis	516	QY	889	GACTTCTTAACCTGGATCAAGGAGGCTGTAATGTCTGTAATAATAGCCGACGCTG	948	Db	517	SerLeuIleAsnThrThrHisGluAsnAlaValAsnValLeuLysAsnThrGlyAsn---	535	QY	949	ACCATCTCCATTGTAGCTGACGTGGCCGGGAGCTGTTATCATCACACCGGAGGGGTG	1009
-----	----	---	-----	----	--------------------------------	-----	----	--	-----	----	---	-----	----	--	-----	----	-----	-------------------------------------	-----	----	---	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	---	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	---	-----	----	-----	-------	-----	----	-----	--	-----	----	-----	-------------------------	-----	----	-----	--	-----	----	-----	---	-----	----	-----	---	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	---	-----	----	-----	--	-----	----	-----	--	-----	----	-----	---	-----	----	-----	--	-----	----	-----	---	-----	----	-----	--	-----	----	-----	---	------

Db	535	-----	-----	-----	536
QY	1009	GCAGAGGCGGCGAGCTGAGCTGCAGCGGCGAGGAGCTTCTCATGCAAGAGCGGCTGC	GC	1068	
Db	536	-----	-----	-----	537
QY	1069	ATGGAGTCCAAACAGATCTCTCAGGAGCAGCAGGAGATGGAGCGCAAGAGGAGAAAGAA	1128		
Db	547	AlaIlePheAsn	-----	550	
QY	1129	ATTCGCCAGAGGCGAGGAGGAAATGAGAGATACCGGAAGGAGTGAACACAGATTGTA	1188		
Db	550	-----	-----	550	
QY	1189	GAGGAGGAAGAGAGTTTAAAGAAGCAATGGGAAGAGACTGGGGCTCAAAGGAACAGCTA	1248		
Db	551	-----	-----	556	
QY	1249	CTCTTGCTAAACCATCATCTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTAT	1308		
Db	557	PheMetProThr	-----	561	
QY	1309	GATCAGGCGAGTGGAAACCTGAGCTCGAGCCCGCAGATGACCTGGATGGAGCAGCGAGGAG	1368		
Db	562	-----	-----	576	
QY	1369	CAGGAGAGCAGAGNTTCCGGAAATATGAGGAAGCGTTTGAC	1425		
Db	577	MetGlySerGlnSerHisLeuSerTyrGlyGlyProLeuAsnThrSerTyrSerSerGln	596		
QY	1426	ACCCGACAGCAGATCATGGGGAGGATGTCGCGCTCTACGCATCAAGAAGGAGGATCC	1485		
Db	597	AlaProIleAlaIleProLeuGluProArgProValGlnLeuValGlyGlyClnAsnGly	616		
QY	1486	TTAGACCTGGCCCTGGAGGCGGTGTGGACTCCCCCATTTGGGAAGGTGGTCTTCTGCT	1545		
Db	617	LeuGlyPheAsnIleValGlyGlyGluAspAsnGlu	634		
QY	1546	GTGTATCAGCGGGAGCTGCTGAGCGCATGTGGCATTTGTAAGAGGGAGCAGATCATG	1605		
Db	635	ValLeuProGlyGlyValAlaAspLeuSerGlyAsnValLysThrGlyAspValLeuLeu	654		
QY	1606	GCAATCAACGCGAAGATTGTACAGACTACACCCCTGGCTGAGCGCTGACGCTCCCTCGAG	1665		
Db	655	GluValAsnGlyValValLeuArgAsnAlaThrHisLysGluAlaAlaGluAlaLeuArg	674		
QY	1666	AAGCCCTGGATCAGGCGGGGACTGGATCGACCTTGTGGTGGCTGCTGCCGCCCAAG	1725		
Db	675	AsnAla	-----	689	
QY	1726	GAGTATCAGCATGAGCTGACCTTCTTGGTGAAGTCCAAAGGGGAACCAATTCACGCG	1785		
Db	690	GluTyrGln	-----	700	
QY	1786	TTAGGAACAGCTGAGCTCCGGCCCCACCTCGTGTGAACACAAAGCCTCGGACGAGC	1839		
Db	701	LeuArgAsnAspValIleAlaGlnSerArgMetGlyThrLeuSerArgLysSer	718		

RESULT 13

T10811

channel associated protein of synapse 2 - rat

C; Species: *Rattus norvegicus* (Norway rat)

```
C:\Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
```

C;Accession: T10811

R; Irie, M.; Hata, Y.; Takai, Y.

submitted to the EMBL Data Library, April 1996

A;Description: Cloning of new isoforms of PSD-95/SAP90 related genes.

A; Reference number: Z17166

A;Accession: T10811

A;Status: preliminary; translated from GB/EMBL/DBBJ

A; Molecule type: DNA

A;Residues: 1-852 <IRI>

A;Cross-references: EMBL:U53368; NID:g1517939; PID:g1517940

C:Genetics:

A:Note: PSD-95/SAP90-related gene 2 (chapsyn isoform2)

C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase homology

F:198-276/Domain: GLGF domain homology <GLG>

F:543-601/Domain: SH3 homology <SH3>

F:663-840/Domain: guanylate kinase homology <GKI>

Alignment Scores:

Pred. No.:	2,15e-06	Length:	852
Score:	225.00	Matches:	144
Percent Similarity:	36.09%	Conservative:	109
Best Local Similarity:	20.54%	Mismatches:	226
Query Match:	5.46%	Indels:	222
DB:	2	Gaps:	33

US-09-502-945-4 (1-2236) x T10811 (1-852)

```

QY 208 GAGGTGGCCCTGCTCGTGGGACACCTGAAGCTGGTCATCAATGAA-----CCGACCCGT 261
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 141 AspGlyArgLeuArgValAsnAspCysIleLeuArgValAsnGluValAspValSerGlu 160
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 262 CTGCCCTCTTTGATCGCATTCGGCGCTGATCCCACTGAAGCACCAGCTGGAATATGAT 321
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 161 ValSerHisSerLysAlaValGlu-----AlaLeuLysGluAlaGlySerIleVal 177
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 322 CAGCTGACCCCGCGCGCTCCAGGAAGCTG-----AAGGAGCTGCGTCTGGAC 369
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 178 ArgLeuTy-ValArgArgArgArgProIleLeuGluThrValValGluIleLysLeuPhe 197
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 370 CGTCTCACCCGAGGCGCTCGGCTGAGTGTGCGTGGTGGC-----CTG 414
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 198 Lys---GlyProLysGlyLeuGlyPheSerIleAlaGlyValGlyValGlyAsnGlnHisIle 216
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 415 GAGTTTGGCTGTGGCTCTCATCTCCACCTCATCAAGGGGCTCAGCAGCAGCGTC 474
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 217 ProGlyAspAsnSerIleTy-ValThrLysIleAlaGlyValGlyValGlyValGlyValGly 236
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 475 GGG---CTCCAGTAGGGAGCAGATCGTCCGGATCAATGATATTCATCTCCTCT 531
    |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 237 GlyArgLeuGlnValGlyAspArgLeuLeuMetValAsnAsnTy-ValSerLeuGluGluVal 256
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 532 ACCCATGAGAGTCAATCACTCATCGA---ACCAAGAAATCTGTCCATCAAGTG 588
    |||||  |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 257 ThrHisGluGluAlaValAlaIleLeuLysAsnThrSerAspValValTy-ValLeuVal 276
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 588 ----- 588
Db 277 GlyLysProThrThrIleTyMetThrAspProTy-ValGlyProProAspIleThrHisSer 296
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 589 -----AGACATC-----GGCCTGATCCCGGTGAAA 615
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 297 TyrSerProProMetGluAsnHisLeuLeuSerGlyAsnAsnGlyThrLeuGluTy-Val 316
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 616 AGCTCT----- 621
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 317 ThrSerLeuProIleSerProGlyArgTy-ValSerProIleProLysHisMetLeuVal 336
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 622 -----CCTGATGAGCCCTCCTGAGTATGCGATGATGCGATGAGTTG 663
    |||  |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 337 GluAspGluTy-ValArgProProGluProVal---TyrSerThrValAsnLysLeuCys 355
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 664 TCGGAATCTGGGGCGTGGCA----- 684
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 356 AspLysProAlaSerProArgHisTy-ValSerProValGluCysAspLysSerPheLeuLeu 375
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 685 -----GGC 687
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 376 SerThrProTy-ValHisLeuGlyLeuLeuProAspSerAspMetThrSerHis 395
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 688 AGCCTGGCTCCCTGGAATCG----- 711
    |||  |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 396 SerGlnHisSerThrAlaThrArgGlnProSerValThrLeuGlnArgAlaIleSerLeu 415
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```

QY 712 GAAACAAAGGAGAGAGAGGTCTTCATCAGCTGTGATGGCTCCCGAGGCGCTTGGCTGCAGC 771
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 416 GluGlyGluProArgLysValValLeuHis---LysGlySerThrGlyLeuGlyPheAsn 434
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 772 ATTTCCAGCGCCCATCAGAGCCTGGCATCTTTTATCAGCCATGTG-----AAA 822
    |||  |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 435 IleValGlyGlyGlu---AspGlyGluGlyPheValSerPheIleLeuAlaGlyGly 453
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 823 CTGGCTCCCTGTCTGTGAGTGGGATGGAGATAGGGACAGATGTGCGAAGTCAAT 882
    |||  |||||  |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 454 ProAlaAspLeuSerGlyGlu-----LeuGlnArgGlyAspGlnIleLeuSerValAsn 471
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 883 GCGTCTGACCTCTTAACCTGGATCACAAGGAGGCTGTAATGTGCTCAAAATAGCGCG 942
    |||||  |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 472 GlyIleAspLeuArgGlyAlaSerHisGlnAlaAlaAlaLeuLysGlyAla--- 490
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 943 AGCTGACCATCTCCATTTAGCTGACGTGGCGGAGCTGTTTCATGCACAGACCGGAG 1002
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 491 GlyGlnThrValThrIleAlaGlnTy-ValGluAspTy-ValAlaArgPheGluAla 510
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1003 CGCTGGCAGAGCGCGCGCAG----- 1023
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 511 LysIleHisAspLeuArgGluGlnMetMetAsnHisSerMetSerSerGlySer 530
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1024 ---CGTAGCTGCGCGCGCAGGCTTCTCATGCAGAGCGGCTGGCGATGGATCCAC 1080
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 531 LeuArgThrAsnGlnLysArgSerLeuTy-ValArgAlaMetPheAspTy-ValSer 550
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1081 AAG-----ATC 1086
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 551 LysAspSerGlyLeuProSerGlnGlyLeuSerPheLysTy-ValGlyAspIleLeuHisVal 570
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1087 CTCAGGAGCAGCAGGATGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1131
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 571 IleAsnAlaSerAspAspGluTrpTrpGlnAlaArgArgValIleLeuAspGlyAspSer 590
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1132 -----GCCAGAGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1170
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 591 GluGluMetGlyValIleProSerLysArgValGluArgLysGluArgAlaArgLeu 610
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1171 GAGATGGAGACAG-----ATTGTAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1212
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 611 LysThrValLysPheAsnAlaLysProGlyValIleAspSerLysGlyAspIleProGly 630
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1213 CAATGGGAGAGAGAGTGGGCTCA-----AAGGACAGCTACTCTTGGCT 1257
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 631 LeuGlyAspAspGlyTy-ValThrAlaThrLeuArgGlyGlnGluAspLeuLeuSer 650
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1258 ---AAACCATCATCTGCTGAG-----GTACACCCAGTACCCCTTCGCAGAGCCA 1302
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 651 TyrGluProValThrArgGlnGluIleAsnTy-ValArgProValIleLeuGlyPro 670
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1303 AAGTATGATCAGGAGTGGAACTGAGCTGAGCCCGCAGATGACCTGGTGGAGGCACG 1362
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 671 MetLysAspArgIleAsnAspLeuIleSerGluPheProAspLysPheGlySerCys 690
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1363 GAGGAGCAGGAGCAGGAGATTTCGGAATATAG---GAAGGCTTTGACCCCTACTCT 1419
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 691 ValProHisThrThrArgProLysArgAspTy-ValValAspGlyArgAspTy-ValHisPhe 710
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1420 ATGTTTCAACCCAGCAGATCATGAGGAGGATGTCGCGCTCCCTACGATCAGAGAGAG 1479
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 711 ValIleSerArgGluGln---MetGluLysAspIleGlnGluHisLysPheIleGluAla 729
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1480 GATCTCTTAGACCTGGCCCTGGAAGC---GGTGTGGACTCCCCCATTTGGGAAGTGTGTC 1536
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 730 GlyGlnTy-ValAspAsnLeuTy-ValGlyThrSerValGlnSer----- 743
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1537 GTTCTGCTGTATGAGCGGGAGCTGCTGAGCGGCAT-----GGT 1578
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 744 ValArgPheValAlaGluArgGly-----LysHisCysIleLeuAspValSerGly 760
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1579 GGCATTGTGAAAGG----- 1593
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```



```

Db 761 AsnAlaIleGlyArgLeuGlnValAlaGlnLeuTyrProIleAlaIlePheIleLysPro 780
QY 1594 -----GACGAGATCATGCAATCAACGCGAGATTGTGACAGACTACACCTGGCT 1644
Db 781 LysSerLeuGluProLeuMetGluMetAsnLysArgLeuThrGluGlnAlaLysLys 800
QY 1645 GAGGCTGACGCTGCCCTGAGAGGCTCGATCAGGCGGGGAGCTGGATCGACCTTGG 1704
Db 801 ThrTyrAspArgAlaIle---LysLeuGluGlnGluPheGlyGlyTyrPheThrAlaIle 819
QY 1705 GTT 1707
Db 820 Val 820
RESULT 14
S71625
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 13 - mouse
N:Alternate names: epidermal growth factor-binding protein; serine proteinase
C:Species: Mus musculus (house mouse)
C:Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 21-Jun-2002
C:Accession: S71625; S67987; I81210; I81209; S40290
R:Chida, D.; Kume, T.; Mukoyama, Y.; Tabata, S.; Nomura, N.; Thomas, M.L.; Watanabe, T.
FEBS Lett. 358, 233-239, 1995
A:Title: Characterization of a protein tyrosine phosphatase (RIP) expressed at a very ea
A:Reference number: S71625; MUID:95145716; PMID:7843407
A:Accession: S71625
A:Molecule type: mRNA
A:Residues: 1-2450 <CHI>
A:Cross-references: EMBL:D83966; NID:g1232103; PIDN:BAAL12158.1; PID:g1232104
A:Experimental source: strain DBA/2; cell line MEL 745A
R:Wolf, B.B.; Brown, M.D.
FEBS Lett. 376, 177-180, 1995
A:Title: Epidermal growth factor-binding protein activates soluble and receptor-bound si
A:Reference number: S67987; MUID:96105375; PMID:7498536
A:Accession: S67987
A:Molecule type: protein
A:Residues: 1098-1102 <WOL>
A:Experimental source: submaxillary glands
R:Sato, T.; Irie, S.; Kitada, S.; Reed, J.C.
Science 268, 411-415, 1995
A:Title: FAP-1: a protein tyrosine phosphatase that associates with Fas.
A:Reference number: I59595; MUID:95232528; PMID:7536343
A:Accession: I81210
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1329-1354, 'K', 1356-1447, 'R', 1449-1454 <RES>
A:Cross-references: GB:L34582; NID:g806297; PIDN:AAC42056.1; PID:g806298
A:Accession: I81209
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1338-1354, 'K', 1356-1447, 'R', 1449-1454 <RE2>
A:Cross-references: GB:L34581; NID:g806295; PIDN:AAC42055.1; PID:g806296
R:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.
submitted to the EMBL Data Library, June 1993
A:Description: Assessment of the expression levels of murine protein-tyrosine phosphatas
A:Reference number: S40280
A:Accession: S40290
A:Molecule type: mRNA
A:Residues: 2266-2372 <HEN>
A:Cross-references: EMBL:D223059; NID:g438155; PIDN:CAA80594.1; PID:g438156
C:Genetics:
A:Gene: Ptpn13
A:Map position: 5
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology;
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; tyros
F:566-860/Domain: protein 4.1 membrane-binding domain homology <B41>
F:1089-1165/Domain: GLGF domain homology <GLG1>
F:1361-1437/Domain: GLGF domain homology <GLG2>
F:1495-1574/Domain: GLGF domain homology <GLG3>
F:1769-1840/Domain: GLGF domain homology <GLG4>
F:1863-1937/Domain: GLGF domain homology <GLG5>
F:2203-2422/Domain: protein-tyrosine-phosphatase homology <PTP1>

```

```

F:2374/Active site: Cys (phosphocysteine intermediate) #status predicted
F:2380/Binding site: substrate phosphate (Arg) #status predicted

Alignment Scores:
Pred. No.: 5,19e-06 Length: 2450
Score: 221.50 Matches: 144
Percent Similarity: 30.88% Conservative: 104
Best Local Similarity: 17.93% Mismatches: 199
Query Match: 5.38% Indels: 356
DB: 2 Gaps: 28

```

US-09-502-945-4 (1-2236) x S71625 (1-2450)

```

QY 308 AGGTGGAATATGATCAGTACCGCCCGCGCTCAGGAAGCTGAAGAGGTGGCTGTGG 367
Db 1070 ArgTrpSerIleValSer--SerProGluArgGluIleThrLeuValAsnLeuLysLysA 1089
QY 368 ACCGTCCTCACCCGAGAGGCTCGGCTGAGTGTGCGGTGGCTGGAGTTTGGC--- 423
Db 1089 spProLysHis-----GlyLeuGlyPheGlnIleIleGlyGlyGlyLysMetGlyArgL 1107
QY 424 ----TGTGGGCTCTTCATCTCCACCTCATCAAGGCGGTGTCAGGCGACAGCGTCGGG- 477
Db 1107 euAspLeuGlyValPheIleSerAlaValThrProGlyGlyProAlaAspLeuAspGlyC 1127
QY 478 --CTCCAGGTAGGGACAGATCGTCCGGATCAATGGATATTCATCTCCTCTGTACCC 535
Db 1127 ysLeuLysProGlyAspArgLeuIleSerValAsnSerValSerLeuGlyValSerH 1147
QY 536 ATGAGGAGGTATCAACCTCATTCGA----- 561
Db 1147 IsHisAlaAlaValAspIleLeuGlnAsnAlaProGluAspValThrLeuValIleSerG 1167
QY 561 ----- 561
Db 1167 lnProLysGluLysProSerLysValProSerThrProValHisPheAlaAsnGlyMetL 1187
QY 562 -----ACCAAGAAACTGTGCCATCAA----- 585
Db 1187 ysSerTyrThrLysLysProAlaTyrMetGlnAspSerAlaMetAspProSerGluAspG 1207
QY 586 -----GTGAGACACATCGGCTGATCCCTGATCCCGTGAAGAGCTCTCGT 625
Db 1207 lnProTrpProArgGlyThrLeuArgHisIle-----ProG 1219
QY 626 ATGAGCCCTCCTACTTGGCAGTATGTGGATCATGTTGTCTCGGAATCTGGGGCGCTGGCA- 684
Db 1219 luSerProPhe-----GlyLeuSerGlyGlyLeuArgG 1230
QY 685 --GGCAGCTGGGCTCCCTCGAATTCGG-----GAAACAAAGG 721
Db 1230 luGlySerLeuSerSerGlnAspSerArgThrGluSerAlaSerLeuSerGlnSerGlnV 1250
QY 722 AGAAGAAGTCTTCATCATCAGCCTGTAGCTCCCGAGG----- 759
Db 1250 alaAsnGlyPhePheAlaSerHisLeuGlyAspArgGlyTrpGlnGluProGlnHisSerS 1270
QY 759 ----- 759
Db 1270 erProSerProSerValThrThrLysValAsnGluLysThrPheSerAspSerAsnArgS 1290
QY 759 ----- 759
Db 1290 erLysAlaLysArgGlyIleSerAspLeuIleGluHisLeuAspCysAlaAspSerA 1310
QY 759 ----- 759
Db 1310 spLysAspAspSerThrTyrThrSerSerGlnAspHisGlnThrSertysGlnGluProS 1330
QY 759 ----- 759
Db 1330 erSerSerLeuSerThrSerAsnLysThrSerPheProThrSerSerAlaSerProProL 1350

```



```
QY 760 -----CTTGCTCAGCA 772
Db 1350 ysProGlyAspThrPheGluValGluLeuAlaLysThrAspGlySerLeuGlyIleSerV 1370
QY 773 TTTCCAGCGC -----CCATCCAGAGCGTGGCATCTTTATCAGCCATGTGAAC 823
Db 1370 alThrGlyGlyValAsnThrSerValArgHisGlyIleThrValLysAlaIleileP 1390
QY 824 CT---GGCTCCCTGTCTGCTGAGTGGGATGGAGATAGGGACCATGTCGAAGTCA 880
Db 1390 roLysGlyAlaAlaGluSerAspGlyArgIleHisLysGlyAspArgValLeuAlaValA 1410
QY 881 ATGGCGTGCACCTTCTCTAACCCTGGATCACAAAGAGGCTGTAATGTGTGAATAATPAGC 940
Db 1410 snGlyValSerLeuGluGlyAlaThrHisLysGlnAlaValGluThrLeuArgAsnThrG 1430
QY 941 GCAGCCTGACCATCTCCATGTTAGCTGCAGCTGCCGCGGAGCTGTTTCATCAGACAGCCGG 1000
Db 1430 lyGlnValVal ----- 1433
QY 1001 AGCGGTGGCAGAGCGCGGACGCTGAGCTGCAGCGGAGGAGCTCTCATGCAGAACG 1060
Db 1434 -----HisLeuLeuLeuGluLys- 1439
QY 1061 GCGTGGATGGAGTCCAAACAGATCCTCCAGAGCAGCAGGAGATGGAGCGCAAGGA 1120
Db 1440 -----GlyGlnValProThrSerArgGluGlnA 1449
QY 1121 GAAAGAAATGGCCAGAGGACAGCAGAGAAATCAGAGATACCCGGAAGGAGATGGAAC 1180
Db 1449 spProAlaGlyProGlnSerProProAspGlnAspAlaGlnArgGlnAlaProGluL 1469
QY 1181 AGATTGTAGAGGAGGAAGAGCAAGTTTAAGAAAGCAATGGGAAGAGACTGGGGCTCAAAG 1240
Db 1469 ysValAlaLysGlnThrProHisValLys -----AspTyr----- 1480
QY 1241 AACAGCTACTCTTGGCTTAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGC 1300
Db 1481 -----SerPheValThrGluAspAsnThrPheGluValLysLeuP 1494
QY 1301 CAAAGTATGATCAGGAGTGGAACCTGAGCTCGAGCCGCGAGATGACCTGGATGGAGGCA 1360
Db 1494 heLysAsnSerSerGlyLeuGlyPheSerPheSerArgGluAspAsnLeu----- 1510
QY 1361 CGGAGGAGCAGGAGAGCAGGATTTCCGAAATATGAGGAAGCTTTGACCCCTACTCTA 1420
Db 1510 ----- 1510
QY 1421 TGTTCACCCAGCAGCAGATCATGGGAGGATGTCGGGCTCTCCAGCATCAAGAGAGG 1480
Db 1511 -----IleProGluGlnIleAsnGlySer -----IleValArgValLysLys--- 1524
QY 1481 GATCCTTAGACTGGCCCTCGAAGCGGTGTGGACTCCCTCCCATGGGAAGTGTGCTTT 1540
Db 1524 ----- 1524
QY 1541 CTGCTGTGTATAGCGGGAGCTGCTGAGCGGCGATGGTGGCATTTGAAAGGGAGCAGA 1600
Db 1525 -----LeuPheProGlyGlnProAlaAlaGluSerGlyLysIleAspValGlyAspValI 1543
QY 1601 TCATGGCAATCAGCGCAAGATTGTGCACACTACACCCCTGGCTGGCTGAGGCTGACGCTGCC 1660
Db 1543 leLeuLysValAsnGlyAlaProLeuLysGlyLeuSerGlnGlnAspValIleSerAlaL 1563
QY 1661 TGCAGAAGCGCTGGAATCAGCGCGGCGGATGATCGACCTGTGTGGTTCGCTGCGCC 1720
Db 1563 euArg-----GlyThrAlaProGluValSerLeuLeuLeuGlyAsp 1577
QY 1721 CA-----AAGAGTATCAGCATGAGCTGACCTCTTCTGCTGAAGTCCAAAA 1765
Db 1577 roAlaProGlyValLeuProGluIleAsp-----ThrThrPheLeu----- 1590
QY 1766 GGGAAACCAAAATTCACGCGTTAGGAACACAGTGAAGTCCGCGCCCTCCCTCGTGAACACAA 1825
```

```
Db 1591 ----AsnProLeuTyrSerProAlaAsnSerPheLeuAsnSerSerLysGluThrSerG 1609
QY 1826 AGCTCGACAGCAGCTTCGAGAGAGCCACATGACACACACACAGCATGGCATCTTGGGACC 1885
Db 1609 InProSerSerValGluGlnGlyAlaSerSerHis-AspAsnGlyValSerGlyLys 1628
QY 1886 TGAATCTATCACCCAGGAATCAAACTCCCTTTGGCCCTGAACAGCGCCAGATGAAGA 1945
Db 1629 ThrLysAsnHis----- 1632
QY 1946 ACAGCTCGGCGACCTTTTGAAGCCCAATGTGGAGGAAGGAGCAG----- 1993
Db 1633 ---CysArgAlaProSerArgGluSerTyrSerAspHisSerGluSerGlyGluAsp 1651
QY 1994 -----CCAGCGCTTTGGGAGAAGATCTCAAGGATCCAGCATCCACACTCTCATTCCT 2038
Db 1652 AspSerValArgAlaProAlaLysMetProAsnValThrArgValAlaAla----- 1668
QY 2039 TTCTCTCGCCCGAGTAATTTGGTCTCTCCAG-----CTTTGGGGG 2080
Db 1669 -----PheProHisGluAlaProArgSerGlnGluSerIleCysAlaMetPheTyr 1686
QY 2081 ACTCTCTCTTGAACCTTAATAAGACCCACTGGAGTCTCTCTCTCTCATCCCTCTCTCT 2140
Db 1687 LeuProArgLysIleProGlyLys-----LeuGluSerGluSerSerHisProPro 1704
QY 2141 CTG 2143
Db 1705 Leu 1705
RESULT 15
A47747
tight junction protein ZO-1 - human
C:Species: Homo sapiens (man)
C>Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
C:Accession: A47747
R:Willott, E.; Balda, M.S.; Fanning, A.S.; Jameson, B.; Van Itallie, C.; Anderson, J.
Proc. Natl. Acad. Sci. U.S.A. 90, 7834-7838, 1993
A:Title: The tight junction protein ZO-1 is homologous to the Drosophila discs-large
A:Reference number: A47747; MUID:93361541; PMID:8395056
A:Accession: A47747
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1736 <ML>
A:Cross-references: GB:L14837; NID:g292937; PID:AAA02891.1; PID:g292938
C:Superfamily: guanylate kinase homology; GLGF domain homology
C:Keywords: alternative splicing; membrane protein; phosphoprotein
F:15-94/Domain: GLGF domain homology <GLGF>
F:181-248/Domain: GLGF domain homology <GLGF2>
F:416-486/Domain: GLGF domain homology <GLGF3>
F:633-782/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 5,33e-06 Length: 1736
Score: 220.50 Matches: 135
Percent Similarity: 36.71% Conservative: 108
Best Local Similarity: 20.39% Mismatches: 244
Query Match: 5.35% Indels: 175
DB: 2 Gaps: 28
US-09-502-945-4 (1-2236) x A47747 (1-1736)
QY 358 GTGCGTCTGACCGCTGTGCACCCCGGAGGCGCTCGAGTGTGGTGGTGGCTGGAG 417
Db 12 ValThrLeuHisArgAlaProGlyPheGlyIleAlaIleSerGlyGlyArgAsp 31
QY 418 -----TTTGGCTGTGG-----CTCTCATCTCCACTCATCAAGCGGT 459
Db 32 AsnProHisPheGlnSerGlyGluThrSerIleValIleSerAspValLeuLysGlyGly 51
QY 460 CAGGCAGACAGCGTGGGCTCCAGGATGAGGAGGAGATGCTCCGCGATCATGATATCC 519
```



```

Db 52 ProAlaGluGly- ---GlnLeuGlnGluAsnAspArgValAlaMetValAsnGlyValSer 70
QY 520 ATCTCTCCTGTTACCCATGAGGAGTCAACCTCATTCGA ---ACCAAGAAACTGTG 576
Db 71 MetAspAsnValGluHisAlaPheAlaValGlnGlnLeuArgLysSerGlyLysAsnAla 90
QY 577 TCCATCAAGTCGAGACACATCGGC ---CTGATCCCGGTGAAAAGCTCTCTCTATGAG 630
Db 91 LysIleThrIleArgArgLysLysValGlnIleProValSerArgProAspProGlu 110
QY 631 CCCCTCACTTGGCAGTATGTGATCAGTTTGTGTGGGAATCTGGGGGTGGAGGCAGC 690
Db 111 ProValSerAspAsnGluAspSerTyrAspGluGluIleHisAspProArgSerGly 130
QY 691 CTGGGCTCCCTGGAAATCGGGAACAAACAGGAG --- 723
Db 131 ArgSerGlyValValAsnArgSerGluLysIleTrpProArgAspArgSerAlaSer 150
QY 723 --- 723
Db 151 ArgGluArgSerLeuSerProArgSerAspArgArgSerValAlaSerSerGlnProAla 170
QY 724 AAGAAAGTCTTCATCAGCCTGGTAGGCTCCCGA ---GGCCTTGGCTGCACGATT 774
Db 171 LysProThrLysValThrLeuValLysSerArgLysAsnGluGluTyrGlyLeuArgLeu 190
QY 775 TCCACGGCCCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAACCTGGCTCCCTG 834
Db 191 AlaSer ---HisIlePheValLysGluIleSerGlnAspSerLeu 204
QY 835 TCTGCTGAGGTGGGA ---TTGGAGATAGGGACAGAGTGTGCAAGTCAATGGCTGCAC 891
Db 205 AlaAlaArgAspGlyAsnIleGlnGluGlyAspValValLeuLysIleAsnGlyThrVal 224
QY 892 TTCTCTAACCTGGATCAACAAGAGGCTGTAAATGTGCTGAAAATAGCGCAGCCTGACC 951
Db 225 ThrGluAsnMetSerLeuThrAspAlaLysThrLeuIleGluArgSerLysGlyLysLeu 244
QY 952 ATCTCCATTGTAGCTGAGCTGGCGGAGCTGTTTC ---ATGACAGAC 996
Db 245 LysMetValValGlnArgAspGluArgAlaThrLeuLeuAsnValProAspLeuSerAsp 264
QY 997 CGGGAGCGCTGGCAGAGCGCGGAGCGT ---GAGCTCAGCGCGGAGCAGG 1044
Db 265 SerIleHisSerAlaAsnAlaSerGluArgAspIleSerGluIleGln --- 281
QY 1045 CTCTCATGACAGAGCGCTGGCATGGATGAGTCCAAACAAGATCCTCCAGGAGCAGCAGG 1104
Db 282 ---SerLeuAlaSerAspHisSerGlyArgSerHisAspArg 294
QY 1105 ATGGAGCGGCAAGAGAGAAAGAAATGCCAGAGAGCAGCAGGAAATGAGAGATAC 1164
Db 295 ProProArgArgSerArgSerArgSerProAspGlnArgSerGluProSerAspHisSer 314
QY 1165 CGGAGGAGATGGAACAG ---ATTGATGAGGAGAGAGAGAAAGTTT 1206
Db 315 ArgHisSerProGlnGlnProSerAsnGlySerLeuArgSerArgAspGluGluArgIle 334
QY 1207 AAGAAGCAATGGGAAGAAGACTGGGGCTCAAAGGAACAGCTACTCTTCCTAAACACATC 1266
Db 335 SerLys ---ProGlyAlaVal 340
QY 1267 ACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCCAAGATGATGATCAGGGAGTGAACCT 1326
Db 341 SerThrProValLysHisAlaAspAspHisThrProLysThrValGluGluVal --- 358
QY 1327 GAGCTGAGCCCGCAGATGACCTGGATGGAGCAGCAGGAGGAGCAGCAGGATTC 1386
Db 359 ---ThrValGluArgAsnGluLysGlnThr 367
QY 1387 CGGAAATATGAGAAAGGCTTTGACCCCTACTCTATGTTTCACCCCGCAGAGCAG --- 1437
Db 368 ProSerLeuProGluProLysProValTyrAlaGlnValGlyAsnGlnMetTrpIleTyr 387

```

```

QY 1438 ---ATCATGGG 1446
Db 388 LeuSerValHisLeuMetValSerTyrLeuIleGlnLeuMetLysMetGlyPheLeuArg 407
QY 1447 AAGATGTCGGCTCCCTACGATCAAGAGGAGGATCTTAGACCTGGCCCTGGAAGGC 1506
Db 408 ProSerMetLysLeuValLysPheArgLysGlyAspSerValGlyLeuArgLeuAlaGly 427
QY 1507 GGTGTGAGTCCCTCATTTGGAGAGTGTCTTCTGTGTGTATGACGGGGGAGCTGCT 1566
Db 428 GlyAsnAsp ---ValGly ---IlePheValAlaGlyValLeuGluAspSerProAla 444
QY 1567 GAGCGCATGTGTGATGTGAAGAGGAGCAGATCATGTGCAATCAACGGCAAGATTGTG 1626
Db 445 AlaLysGlu ---GlyLeuGluGluGlyAspGlnIleLeuArgValAsnValAspPhe 463
QY 1627 ACAGACTACACCTGGTGGCTGAGGCTGACGCTGCCCTGCAGAGGCTGGAATCAGGGCGG 1686
Db 464 ThrAsnIleIleArgGluGluAlaValLeuPheLeu --- 475
QY 1687 GACTGGTACACCTGTGTGTGCTGCTGCCCCCAAGGAGTATGACGATGAGCTGACC 1746
Db 476 ---LeuAspLeu ---ProLys ---GlyGluGluValThr 485
QY 1747 TCTGTGTGAAGTCCAAAGAGGGAAACCAATTCACGCGTTAGGAAACAGTGAGCTCCGG 1806
Db 486 IleLeuAlaGlnLysLysLys ---AspValTyr 495
QY 1807 CCCACCTCTGACACACAAAGCCCTCGGACCAAGCTTCAG ---AGAGCCACATGACACAC 1863
Db 496 ArgArgIleValGluSerAspValGlyAspSerPheTyrIleArgThrHisPheGluTyr 515
QY 1864 ACCAGATGGCATCTTGGGACCTGAATCTATCACCCAGGAATCTCAAACTCCCTTTGGCC 1923
Db 516 GluLysGluSerProTyrGly ---LeuSer 524
QY 1924 CTGAACAGGCGCCAGATAAGAAACAGCTCGGGCCACTTTTGAAGCCCAATGTGGAGGA 1983
Db 525 PheAsnLysGlyGluValPheArgAlaValAspThrLeuTyrAsnGlyLysLeu ---Gly 543
QY 1984 AAGGAGCAGCCACCCCTTTGGGAGAGATCTCAAGGATCCAGACTCTCATTCTCTTCCT 2043
Db 544 SerTrpLeuAlaIleArgIleGlyLysAsnHisLysGluValGluArgGlyIleIlePro 563
QY 2044 ---CTGGCCAGTGAATTTGGTCTCTCCCA ---GCTTTG 2076
Db 564 AsnLysAsnArgAlaGluGlnLeuAlaSerValGlnTyrThrLeuProLysThrAlaGly 583
QY 2077 GGGGAC 2082
Db 584 GlyAsp 585

```

Search completed: March 21, 2003, 13:05:51
Job time : 77.8301 secs

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	2823	68.5	552	1	A175_HUMAN	Q9Y6n9	homo sapien
2	275.5	6.7	933	1	Z03_HUMAN	Q95049	homo sapien
3	268.5	6.5	905	1	Z03_MOUSE	Q9QXV1	mus musculus
4	263.5	6.4	767	1	D1G4_HUMAN	P78352	homo sapien
5	261.5	6.3	724	1	D1G4_RAT	P31016	rattus norv
6	261.5	6.3	1745	1	Z01_MOUSE	P39447	mus musculus
7	258.5	6.3	724	1	D1G4_MOUSE	Q62108	mus musculus
8	249.5	6.1	898	1	Z03_CANFA	O62683	canis famill
9	244.5	5.9	904	1	D1G1_HUMAN	Q12959	homo sapien
10	243	5.9	849	1	D1G3_MOUSE	P70175	mus musculus
11	243	5.9	849	1	D1G3_RAT	Q62936	rattus norv
12	240.5	5.8	911	1	D1G1_RAT	Q62936	rattus norv
13	234.5	5.7	817	1	D1G3_HUMAN	Q92796	homo sapien
14	230	5.6	870	1	D1G2_HUMAN	Q15700	homo sapien
15	229	5.6	852	1	D1G2_RAT	Q63622	rattus norv
16	220.5	5.4	1736	1	Z01_MOUSE	Q07157	homo sapien
17	195.5	4.7	2485	1	PTND_HUMAN	Q12923	homo sapien
18	193.5	4.7	2161	1	SHK1_HUMAN	Q9Y566	homo sapien

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----

DR EMBL; AF039700; AAC18049.1; ALT_FRAME.
DR EMBL; AF039699; AAC18048.1; -
DR EMBL; AB006955; BA81739.1; -
DR EMBL; AB018687; BA81740.1; -
DR HSSP; P29476; IOAV.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 6.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS50106; PDZ; 3.
KW Antigen; Repeat; Alternative splicing; Polymorphism.
FT DOMAIN 87 169 PDZ 1.
FT DOMAIN 211 293 PDZ 2.
FT DOMAIN 452 537 PDZ 3.
FT VARSPLIT 1 31 MISSING (IN ISOFORM 2).
FT VARSPLIT 404 552 MISSING (IN ISOFORM 3).
FT VARIANT 519 519 D -> E (IN DBSNP:1064074).
/FTID=VAR_012320.
FT CONFLICT 103 103 R -> S (IN REF. 2; BAA81739).
FT CONFLICT 280 280 N -> S (IN REF. 3).
FT CONFLICT 305 305 A -> T (IN REF. 2; BAA81739).
SQ SEQUENCE 552 AA; 62223 MW; 7B405B02CACE8241 CRC64;

Alignment Scores:

Pred. No.: 1,15e-145 Length: 552
Score: 2823.00 Matches: 551
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 68.52% Indels: 0
DB: 1 Gaps: 0

US-09-502-945-4 (1-2236) x AI75_HUMAN (1-552)

QY 97 ATGGACCGAAAGTGGCCGAGAAATTCGGCATAGAGTGGATTTCTGATTGAAATGAT 156
DB 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
QY 157 GCAGAGAAGGACTATCTCTATGATGTGTCGAATGTACCAACAGACCATGAGCTGGCC 216
DB 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
QY 217 GTGCTGTGGGAGACCTGAAGCTGTGCATCAATGAACCCAGCCGCTGCTCTGTTGAT 276
DB 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
QY 277 GCATTGCGCCGCTGATCCCTGAAGCAGCAGGTGGATATGATCAGTACCCCGCGG 336
DB 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
QY 337 CGCTCCAGGAAGCTGAAGAGGTGGCTCTGGACCGCTGTGCACCCCGAAGCCCTCGCCCTG 396
DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGluLeuGlyLeu 100
QY 397 AGTGTGCTGGTGGCTGGAGTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGAGC 456
DB 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
QY 457 GGTCAAGCAGACGCTGGGCTCCAGGTAGGAGGAGAGTCCCGGATCAATGATGATAT 516
DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
QY 517 TCATCTCTCTCTGTACCCATCAGGAGGTCAATCAACTCATTCGAACCAAGAAACTGTG 576
DB 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
QY 577 TCCATCAAAAGTGAGACATCGCCCTGATCCCGGTGAAAGAGCTCTCTGATGAGCCCTC 636

DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
QY 637 ACTTGGCAGTATCGATCAGTTGTGTCGGAATCTGGGGCGTGGAGCAGCAGCTGGGC 636
DB 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200
QY 697 TCCCTCGAAATCGGAAACAAGAGAGAGAGTCTTCATCAGCCTGGTAGCTCCCGA 756
DB 201 SerProGlyAsnArgGluAsnLysGluLysLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCTTGGCTGCAGCATTTCCAGCGCCCATCCAGAAGCCTGSCATCTTTTATCAGCAT 816
DB 221 GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTGAACCTGCTGCTCTGTGAGTGGGATGGAGATAGGAGCAGCAGATTTGTGAA 876
DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCAATGGCTGACTTCTTAACCTGGATCACAAGAGGCTGTAAATGTGCTGAAAT 936
DB 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
QY 937 AGCGCAGCCTGACCATCTCCATTGTAGTGCAGCTGGCGGAGCTGTTTCATGACAGAC 996
DB 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGAGCGCTGGCAGAGCGCGCAGCTGAGCTGCAGCGGAGGAGCTTCTCATGCAG 1056
DB 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
QY 1057 AAGCGCTGGGATGGAGTCCACAAGATCTCTCAGAGCAGCAGGAGATGGAGCGCAA 1116
DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnMetGluArgGln 340
QY 1117 AGGAGAAAGAAATGCCAGAGCAGCAGAGAAATCAGAGATACCGGAAGAGAGATG 1176
DB 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluLysGluAsnGluArgTyrArgLysGluMet 360
QY 1177 GAACAGATTCTAGAGGAGAGAGAGTAAAGCAATGGGAAGAGACTGGGGCTCA 1236
DB 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380
QY 1237 AAGGAACAGCTACTTTCCTAAACCATCACTCTCAGGTACACCCAGTACCCCTTCGC 1296
DB 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCAAAGTATGATCAGGAGTGAACCTGAGCTCCAGCCCGCAGATGACCTGGATGA 1356
DB 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspAspLeuAspGly 420
QY 1357 GGCACGAGGAGCAGGAGCAGGATTCGGAAATATCAGGAAGCTTTGACCCCTAC 1416
DB 421 GlyThrGluGluGlnGlyGluAspPheArgLysTyrGluGluGlyPheAspProTyr 440
QY 1417 TCTATGTTACCCAGACAGCAGATCATGGGAAGGATGTCCGGCTCTCAGCATCAAGAAG 1476
DB 441 SerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLys 460
QY 1477 CAGGATCTCTAGACCTGGCCCTGGAAGCGGTGTGACTCCCTCCATTGGGAAGTGGTC 1536
DB 461 GluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProIleGlyLysValVal 480
QY 1537 GTTCTGTGTGTATGAGCGGGAGCTGCTGAGCGGATGCTGGCATTTGTGAAAGGGAC 1596
DB 481 ValSerAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAsp 500
QY 1597 GAGATCATGCAATCAACGCAAGATTGTACAGACTACACCTGCGCTGAGCTGAGCTGAGCT 1656
DB 501 GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAla 520
QY 1657 GCCTGCAAGAGGCTGGAATCAGGGGGGAGCTGGATCGACCTTGTGGTGGCTGCTGC 1716


```

Db 521 AlaLeuGlnLysAlaTrpAsnGlnGlyGlyAspTrpIleAspLeuValAlaValCys 540
QY 1717 CCCCCAACGAGTATGACGATGAGTGACCTTC 1749
Db 541 ProProLysGluTyrAspAspGluLeuThrPhe 551
RESULT 2
Z03_HUMAN STANDARD; PRT; 933 AA.
AC O95049;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
DE occludens 3 protein) (Tight junction protein 3).
GN TJP3 OR Z03.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Brower A., Gordon L., Dias J., Ramirez M.,
RA Stillwagen S., Phan H., Velasco N., Do L., Regala W., Terry A.,
RA Ganes J., Danganan L., Erier A., Christensen M., Georgescu A.,
RA Avila J., Liu S., Attix C., Andreise T., Trankheim M.,
RA Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R.,
RA Thomas P., Quan G., Krommiller B., Arellano A., Sanders C., Ow D.,
RA Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.,
RT "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
RT serine protease gene cluster."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: INTERACTS WITH OCLUDIN, CLAUDINS AND ZO-1 (BY
CC SIMILARITY).
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC005954; AAC72274.1; ALT_INIT.
DR HSP; P31016; 1BFE.
DR Genew; HGNC:11829; TJP3.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR SMART; SM00072; GuK; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50106; PDZ; 3.
DR TJT junction; SH3 domain; Repeat; Membrane.
KW DOMAIN 11 93
FT DOMAIN 195 272 PDZ 1.
FT DOMAIN 394 460 PDZ 2.
FT DOMAIN 489 563 PDZ 3.
FT DOMAIN 675 775 GUANYLATE_KINASE.
SQ SEQUENCE 933 AA; 102800 MW; 680298CFD0615B47 CRC64;
Alignment Scores: 8.16e-08 Length: 933
Pred. No.:

```

```

Score: 275.50 Matches: 142
Percent Similarity: 35.97% Conservative: 90
Best Local Similarity: 22.02% Mismatches: 210
Query Match: 6.69% Indels: 203
DB: 1 Gaps: 28

US-09-502-945-4 (1-2236) x Z03_HUMAN (1-933)
QY 385 GGCTCTGGCTGAGTGTGGCTGGCTGGAGTTT---GGCTGTGGCTCTTCATCTCC 441
Db 21 GlyPheGlyIleAlaIleSerGlyArgAspArgProGlyGlySerMetValValSer 40
QY 442 CACCTCATCAAGCGGTCCAGCAGACAGCGTCGGGCTCCAGGTAGGAGGACGATCTGTC 501
Db 41 AspValValProGlyGlyProAlaGluGly---ArgLeuGlnThrGlyAspHisIleVal 59
QY 502 CGGATCAATGATATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 561
Db 60 MetValAsnGlyValSerMetGluAsnAlaThrSerAlaPheAlaIleGlnIleLeuLys 79
QY 562 ACC---AAGAAACTGTGTCTCATCAAAAGTG-----AGACACATCGGCTGATCCCC 609
Db 80 ThrCysThrLysMetAlaAsnIleThrValLysArgProArgArgIleHisLeuProAla 99
QY 610 GTGAAAGCTCTCTCT---GATGAGCCCTC 636
Db 100 ThrLysAlaSerProSerSerProGlyArgGlnAspSerAspGluAspGlyProGln 119
QY 637 ACTGGCAGTAGTGGATCAG-----TTTGTCGCGAATCTGGG 675
Db 120 ArgValGluGluValAspGlnGlyArgGlyTyrAspGlyAspSerSerGlySerGly 139
QY 676 -----GCGTCGCGAGGAGCGCTGGGCTCC 699
Db 140 ArgSerTrpAspGluArgSerArgArgProArgProGlyArgArgGlyArgAlaGlySer 159
QY 700 CTGGAATCGGAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 759
Db 160 HisGlyArgArgSerProGlyGly-----GlySerGluAla 171
QY 760 CTGCTGCATCATTTCCAGCGGC-----CCCATCCAG-----AAG 795
Db 172 AsnGlyLeuAlaLeuValSerGlyPheLysArgLeuProArgGlnAspValGlnMetLys 191
QY 796 CCT----- 798
Db 192 ProValLysSerValLeuValLysArgArgAspSerGluGluPheGlyValLysLeuGly 211
QY 799 ---GGCATCTTTATCAGCATGTGAAACCTGGCTCCCTGTCTGCT---GAGGTGGATTG 852
Db 212 SerGlnIlePheIleLysHisIleThrAspSerGlyLeuAlaAlaArgHisArgGlyLeu 231
QY 853 GAGATAGGCGACGAGATGTTCGAAGTCAATGGCTGCGACTTCTTAACCTCGATCACAAG 912
Db 232 GlnGluGlyAspLeuIleGlnIleAsnGlyValSerSerGlnAsnLeuSerLeuAsn 251
QY 913 GAGCGTGAATGTCTGAAAAATAGCCGCGACCTGACCATCTCCATTGTAGTGCAGCT 972
Db 252 AspThrArgArgLeuIleGluLysSerGluGly---LysLeuSerLeuValLeuArg 270
QY 973 GCGCGGAGCTGTCATCAGACAGCGGAGCGGTGGCAGAGCGCGCGCGCGCGAGCTG 1032
Db 271 AspArgGlyGlnPheLeuValAsnIleProProAlaValSerAspSerAspSerPro 290
QY 1033 CAGCGGAGGAGCTTCTCATCAGAGCGGTGGCGATGAGTCCACACAGATCCTCCAG 1092
Db 291 LeuGluGluGlyValThrMetAlaAspGluMetSerSerProProAlaAspIleSerAsp 310
QY 1093 GAGCAGCAGGAGATGGAGCGG----- 1113
Db 311 LeuAlaSerGluLeuSerGlnAlaProProSerHisIleProProProProArgHisAla 330
QY 1114 CAAGGAGAAAAGAAATTTGCCAG---AAGCAGCAGAGGAAAATGAGAGATACCGGAAG 1170

```



```

Db 60 MetValAsnGlyValSerValGluAsnValThrSerAlaPheAlaIleGlnIleLeuLys 79
QY 562 ACC---AGAAAACGTGTCCATCAAGTG-----AGACATCGCCTGATCCCC 609
Db 80 ThrCysThrLysThrAlaAsnValThrValLysArgProArgArgValGlnLeuProAla 99
QY 610 GTGAAGACTCTCCCTGATGAGCCCTCACTTGGCAGTATGTGATGATGTCGCGAA 669
Db 100 ThrLysAlaSerProAlaSer-----GlyHisGlnLeuSerAspGlnGluAlaAsp 117
QY 670 TCTGGG----- 675
Db 118 HisGlyArgGlyTyrGluGlyAspSerSerGlySerGlyArgSerTrpClyGluArg 137
QY 676 -----GGCTGGCAGGCAGCCCTGGCTCCCTCGAATCGGAAAC 717
Db 138 SerArgArgSerArgAlaGlyArgGlyArgValGlySerHisGlyArgArgSerSer 157
QY 718 -----AAG 720
Db 158 GlyGlyGlySerGluAlaAsnGlyLeuAspLeuValSerGlyTyrLysArgLeuProLys 177
QY 721 GAG-----AAGAGGTCTTCATCAGCCTGGTAGGCTCCCGAGGC 759
Db 178 GlnAspValLeuMetArgProLeuLysSerValLeuValLysArgArgAsnSerGlu 197
QY 760 CTTGGCTGCAGCATTTCCAGCGCCCTCCAGAGCCTGGATCTTTATCAGCATGTG 819
Db 198 PheGlyValLysLeuGlySer-----GlnIlePheIleLysHisIle 211
QY 820 AAACCTGGCTCCTCTCTGCT---GAGGTGGATTGGAGATAGGAGACACAGATTCGCA 876
Db 212 ThrGluSerGlyLeuAlaAlaArgAsnHisGlyLeuGlnGluGlyAspLeuIleGln 231
QY 877 GTCATATGGCTGCTCTCTTAACCTGGATCAACAAGGAGCTGTAATGTGCTGAAAAT 936
Db 232 IleAsnGlyValSerSerAlaAsnLeuSerLeuSerAspThrArgArgLeuIleGluLys 251
QY 937 AGCCGC---AGCCTACCATCTCCATTGTAGCTGCAGCTGGCGGGAGCTG----- 984
Db 252 SerGluGlyGluLeuThrLeuLeuValLeuArgAspSerGlyGlnPheLeuValAsnIle 271
QY 985 -----TTCATGACAGAC-----CGGAG 1002
Db 272 ProProAlaValSerAspSerAspSerSerLeuMetGluAspIleSerAspLeuThrSer 291
QY 1003 CGGTGGCAGAGCGCGCAGCTGAGCTGAGCGGCGAGGAGCTTCTCATCAGAAGCGG 1062
Db 292 GluLeuSerGlnAlaProSerHisValProProProProLeuLysGlyGln---Arg 310
QY 1063 CTGGCGATGGAGTCCAAACAGATCTCCAGGAGCAGCAGGAGATGGAGCGCAAGGAGA 1122
Db 311 SerProGluAspSerGlnThrAspSerProValGluThrProGlnProArgArgGlu 330
QY 1123 AAAGAATATGGCCAGAGCGCAGCAGAGAAATGAGACATACCCGGAAGAGATGAACAG 1182
Db 331 ArgSerValAsnSerArgAlaIleAlaGluProGluSer----- 343
QY 1193 ATTGTAGAGGAAGAGAAGTTTAAGAAGCAATGGGAAGAGACTGGGGCTCAAGAA 1242
Db 343 ----- 343
QY 1243 CAGCTACTCTTGCCTAAACCATCACTCTGAGGTACACCCAGTACCCCTTCGCAAGCCA 1302
Db 344 -----ProGlyGluSerArgTyrAspIleTyrArgValProSerArgGlnSer 359
QY 1303 AAGTATGATCAGGAGTGGACCTCAGCTCCGAGCCGAGATGACCTGGATGGAGGCAG 1362
Db 360 LeuGluAspArgGlyTyrSerPro----- 367
QY 1363 GAGGAGCAGGAGAGCAGGATTTCCGGAATATGAGGAAGGCTTTGACCCCTACTCTATG 1422
Db 367 ----- 367

```

```

QY 1423 TTCACCCAGACAGATCATGGGAAGGATGTCGGGCTCTCATGCATCAAGAAGAGGGA 1482
Db 368 -----AspThrArgValValSerPheProLysGlyAla 378
QY 1483 TCCTTAGACCTGGCCCTGGAAAGCGGTGTGACACTCCCCCATTTGGGAAGGTGGTCTTCT 1542
Db 379 SerIleGlyLeuArgLeuAlaGlyGlyAsnAsp-----ValGly---IlePheValSer 395
QY 1543 GCTGTGTATGAGCGGGAGCTGCTGAGCGCATGGTGCATTTGTAAGGGGACGAGATC 1602
Db 396 GlyVal---GlnAlaGlySerProAlaAspGlyGlnGlnIleGlnGluGlyAspGluIle 414
QY 1603 ATGCAATCAACGCAAGATTTGTACAGACTACACCTGGCTGGCTGAGGTGAGCTGCCCTG 1662
Db 415 LeuGlnValAsnGlyMetProPheArgAsnLeuThrArgGluGluAla----- 430
QY 1663 CAGAAGCCTGGATCAGGCGGGAGCTGATGACCTTGTGGTTCGCTGCGCTGCCCCCA 1722
Db 431 -----ValGlnPheLeuLeuGlyLeu---ProPro 439
QY 1723 AAGGATATGACGATGAGCTGACCTTCTGTGAAGTCCAAAGCGGAACCAATTCAC 1782
Db 440 GlyGlu-----AspMetGluLeuValThrIleSerLysThrGlyHisSerLeuArg 456
QY 1783 CGCTTAGAAACAGTGTGCTCCGCGCCACCTCGTGAACACAAAGCCTCGGACAGCCTT 1842
Db 457 ArgTrpSerGlnSerArgVal----- 463
QY 1843 GAGAGAGCCACATGACACACAGATGGCATCTCTTGGGACCTGAATCATCATCACCAGG 1902
Db 464 -----GlyAspSerPheTyrIleArgThrHis---PheGluLeuGluProSerPro--- 479
QY 1903 AATCTAAACTCCCTTTGGCCTGAACACAGGCGCAGATA----- 1941
Db 480 -----ProTyrGlyLeuGlyPheThrArgGlyAspValPheHisValValAspThrLeu 497
QY 1942 -----AGGAACACCTCGGCGCACTTTTGAAGCCCAATGTGGAGGAAGGAGCAGCC 1995
Db 498 TyrProGlySerGlyProGlyHisGlyHisSerHis---GlyGlyLeuTrpLeuAla 516
QY 1996 AGCGCTTTGGGAGAGATCTCAAGATCCAGACTCTCATCTCTTCCTTCCT 2043
Db 517 AlaArgMetGlyArgAspLeuArgGlnGlnGluArgGlyValIlePro 532

```

RESULT 4

```

DLG4_HUMAN
ID DLG4_HUMAN STANDARD; PRT; 767 AA.
AC P78352; Q92941; Q9UKK8;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic density protein 95 (PSD-95) (Discs, large homolog 4)
DE Postsynaptic density-95).
GN DLG4 OR PSD95.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=9743282; PubMed=9286702;
RA Stathakis D.G., Hoover K.B., You Z., Bryant P.J.;
RT "Human postsynaptic density-95 (PSD95): location of the gene (DLG4)
and possible function in nonneural as well as in neural tissues.";
RL Genomics 44:71-82(1997).
RN [2]
RP REVISIONS.
RC TISSUE=Mammary gland;
RA Stathakis D.G., Hoover K.H., You Z., Bryant P.J.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]

```



```

QY 1216 TGGGAA-----GAGACTGGGGCTCA-----AAGGAA 1242
Db 542 TrpSerArgLeuLysAlaLysAspTrpGlySerSerGlySerGlnGlyArgGluAsp 561
QY 1243 CAGTACTCTTGCCTAAACCATCACT---GCTGAGGTACAC-----CCAGTACCC 1290
Db 562 SerValLeuSerTyrGluThrValThrGlnMetGluValHisTyrAlaArgProIlele 581
QY 1291 CTTCGCAAGCCAAAGTATGATCAGGAGTGGAACTGAGCTGAGCTGAGCCCGCAGATGACCTG 1350
Db 582 IleLeuGlyProThrLysAspArgAlaAsnAspLeuLeuSerGluPheProAspLys 601
QY 1351 GATGGAGCAGGAGGAGCAGGAGCAGGAGTATTCGGAATATGAGAGGCTTTTCAC 1410
Db 602 PheGlySerCysValProHisThrThrArgProLysArgGluTyrGluIleAspGlyArg 621
QY 1411 CCTTACTTATGTTTACCCAGAGCAGATCATGGGGAAGGATGCTCGGCTCTAGCGCATC 1470
Db 622 AspTyrHisPheValSerSerArgGluLysMetGluLysAspIleGlnAlaHisLysPhe 641
QY 1471 AAGAGGAGGATCTTACACTGCGCTCGAAGCGGTGGAGCTCCGCCATT----- 1524
Db 642 IleGluAlaGlyGlnTyrAsnSerHisLeuTyrGlyThrSerValGlnSerValArgGlu 661
QY 1525 -----GGGAAG-----GTGGTCTGTTCTGCTGTGTATGAGCGG 1557
Db 662 ValAlaGluGlnGlyLysHisCysIleLeuAspValSerAlaAsnAlaValArgLeu 681
QY 1558 GGAGCTGTGAGCGCGCTGGTGGCATTTGTG-----AAAGGGCAGGAGATC 1602
Db 682 GlnAlaAlaHisLeuHisProIleAlaIlePheIleArgProArgSerLeuGluAsnVal 701
QY 1603 ATGGCAATCAGCGCAAGATGTTGACAGACTACACCTCGGTGAGCTGAGCTGCC 1659
Db 702 LeuGluIleAsnLysArgIleThrGluGlnAlaArgLysAlaPheAspArgAla 720

RESULT 5
DLG4_RAT STANDARD; PRT; 724 AA.
AC P31016; P97631;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)
DE (Synapse-associated protein 90) (Discs, large homolog 4).
GN DLG4 OR PSD95.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=93040233; PubMed=1419001;
RA Cho K.-O., Hunt C.A., Kennedy M.B.;
RT "The rat brain postsynaptic density fraction contains a homolog of
RT the Drosophila discs-large tumor suppressor protein.";
RL Neuron 9:929-942(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=93186749; PubMed=7680343;
RA Kistner U., Wenzel B.M., Veh R.W., Cases-Langhoff C., Garner A.M.,
RA Appettau U., Voss B., Gundelfinger E.D., Garner C.C.;
RT "SAP90, a rat presynaptic protein related to the product of the
RT Drosophila tumor suppressor gene dlg-A.";
RL J. Biol. Chem. 268:4580-4583(1993).
RN [3]
RP SEQUENCE OF 566-625 FROM N.A.
RC STRAIN=Wistar Kyoto; TISSUE=Vascular smooth muscle;
RA Adams L.D., Werny I., Schwartz S.M.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [4]

```

```

RP INTERACTION WITH DLGAP1 AND SHANK PROTEINS.
RX PubMed=10527873;
RA Boeckers T.M., Winter C., Smalla K.-H., Kreutz M.R., Bockmann J.,
RA Seidenbecher C., Garner C.C., Gundelfinger E.D.;
RT "Proline-rich synapse-associated proteins prosapi and prosap2 interact
RT with synaptic proteins of the SAPAP/GKAP family.";
RL Biochem. Biophys. Res. Commun. 264:247-252(1999).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS) OF 302-402.
RX MEDLINE=96270509; PubMed=8674113;
RA Doyle D.A., Lee A., Lewis J., Kim E., Sheng M., Mackinnon R.;
RT "Crystal structures of a complexed and peptide-free membrane protein-
RT binding domain: molecular basis of peptide recognition by PDZ.";
RL Cell 85:1067-1076(1996).
CC -|- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
CC -|- SUBUNIT: Interacts with DLGAP1/GKAP. Is part of a
CC complex with DLGAP1/GKAP, SHANK1 SHANK2 or SHANK3.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic. Concentrated at synaptic
CC junctions primarily on the presynaptic side. Also found in
CC postsynaptic density of neuronal cells.
CC -|- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.
CC -|- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96853; AAA1971.1; -
DR EMBL; X66474; CAA47103.1; -
DR EMBL; U77090; AAB38270.1; -
DR PIR; S26407; S26407.
DR PIR; JH0800; JH0800.
DR PDB; 1BE9; 21-OCT-98.
DR PDB; 1BE9; 21-OCT-98.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS00552; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain; Repeat; 3D-structure.
FT DOMAIN 65 151 PDZ 1.
FT DOMAIN 160 246 PDZ 2.
FT DOMAIN 313 393 PDZ 3.
FT DOMAIN 428 498 SH3.
FT DOMAIN 534 724 GUANYLATE_KINASE.
FT CONFLICT 61 61 M -> L (IN REF. 2).
FT CONFLICT 78 78 S -> T (IN REF. 2).
FT CONFLICT 177 182 GVGNOH -> ALGTSI (IN REF. 2).
FT CONFLICT 200 200 A -> G (IN REF. 2).
FT CONFLICT 254 254 S -> T (IN REF. 2).
FT CONFLICT 540 555 LGPTKDRANDLLSEF -> SLDPKTYPTMTIFPSS (IN REF. 2).
FT CONFLICT 623 625 GRH -> RDO (IN REF. 3).
SQ SEQUENCE 724 AA; 80465 MW; 7922D4E8E0F9AD85 CRC64;

```

Alignment Scores:

Pred. No.:	4.47e-07	Length:	724
Score:	261.50	Matches:	125
Percent Similarity:	40.45%	Conservative:	93
Best Local Similarity:	23.19%	Mismatches:	206
Query Match:	6.35%	Indels:	115
DB:	1	Gaps:	24

US-09-502-945-4 (1-2236) x DLG4_RAT (1-724)

QY	340	TCCAGGAAGCTGAAGGAGTGGCTGTGGACCGCTGCACCOCGAAAGCCCTCGGCCCTGAGT	399
Db	155	AAGLulysValMetGlulileLysLeuIleLys---GlyProLysGlyLeuGlyPheSer	173
QY	400	GTGGTGCTGGC-----CTGGAGTTGGCTGTGGCTTCATCATCTCCCAC	444
Db	174	IleAlaGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrrValThrLys	193
QY	445	CTCATCAAGCGCGTCAAGCACACGCTCGGG---CTCCAGGTAGGGGACGAGATCGTC	501
Db	194	IleIleGluGlyAlaAlaHisLysaspGlyArgLeuGlnIleGlyAspLysIleLeu	213
QY	502	CGATCATAGGATATTCCATCTCTCTCTGTCACCATGAGGAGTCATCACTCATTCGA	561
Db	214	AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaLeuLys	233
QY	562	--ACCAAGAAAACACTGTCCAATCAAAGTG-----AGACACATCGCGCTGATCCCC	609
Db	234	AsnThrTyraSpValValltyrLeuLysValAlaLysProSerAsnAlaTyrrLeuSerAsp	253
QY	610	GTAAAAAGCTCTCCTGATGAGCCCTCACTGG---CAGTATGTGGATCAGTTTGTGTCG	666
Db	254	SerTyraLaProProaspIleThrThrSertyrSerGlnHisLeuAspAsnGluIleSer	273
QY	667	GAATCTGGGGCGTCGCAGCACCTCTGGC-----	696
Db	274	Hisser-----SerTyrrLeuGlyThrAspTyrrProThrAlaMetThrProThr	289
QY	697	-----TCCCTGTGAAATCGGGAA-----AACAAAG	720
Db	290	SerProArgArgTyrrSerProValAlaLysAspLeuGlyGluAspIleProArg	309
QY	721	GAGAAGAAGCTCTTCATCAGCCTGGTAGCCTCCGAGGCGTTGGCTGCAGCATTTCCAGC	780
Db	310	GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly	329
QY	781	GGCCCCATCCAGAAGCTGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC	831
Db	330	GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp	348
QY	832	CTGTCTCTGAGTGGGATTGGAGATAGGGGACCAAGATTGTCAAGTCAATGGCTGCAC	891
Db	349	LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValasnGlyValasp	366
QY	892	TTCTCTAACCTGGATCACAAGGAGGCTGTAATGTGCTGAAAAATAGCCGAGCCTGACC	951
Db	367	LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr	385
QY	952	ATCTCCATTTAGCTGAGCTGCCCGGGAGCTGTTTCATGACAGACGGGAGCGCTGCCA	1011
Db	386	ValthrIleIleAlaGln-----TyrLysProGluGlyTyrrSerArgPhe	400
QY	1012	GAGCGCGGCGAGCTGAGTGTGACGGCAGGAGCTTCTCATCAGAACGGCTGCGCATG	1071
Db	401	GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer	418
QY	1072	GAGTCCCAACAGATCTCCAGGAGCAGCAGGAG-----ATCGAG	1110
Db	419	GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrrIleArgAlaLeuPheAsp	438
QY	1111	CGGCAAGAGGAGAAAGAAATTGCCAGAGGCGCAGAGGAGAAATGACAGATACCGGAG	1170
Db	439	TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp	458


```

Db 416 -----GlylleLeuArgProSerMetLys 423
QY 1459 CTCCTACGATCAAGAAGGAGGATCTTAGACCTGGCCCTGGAAGGCGGTGTGACATCC 1518
    |||:||||| |||: ||| ||| ||| ||| |||
Db 424 LeuValLysPheArgLysGlyAspSerValGlyLeuArgLeuAlaGlyGlyAsnAsp--- 442
QY 1519 CCCATTGGGAAGTGGTCTGCTCTGCTGTATGAGCGGGGAGCTGCTGACGGCATGGT 1578
    |||:||||| |||: ||| ||| ||| ||| |||
Db 443 ---ValGly---IlePheValAlaGlyValLeuGluAspSerProAlaAlaLysGlu--- 459
QY 1579 GGCATTGTGAAGGGGAGGATGATGCAATCAACGGCAAGATGTCACACAGATACACC 1638
    |||:||||| |||: ||| ||| ||| ||| |||
Db 460 GlyLeuGluGluGlyAspGlnIleLeuArgValAsnAsnValAspPheThrAsnIleile 479
QY 1639 CTGGCTGAGGTGAGCGCTGCCCTCGAGAGCGCTGGAATCAATGAGGGGCGGATGCAC 1698
    |||:||||| |||: ||| ||| ||| ||| |||
Db 480 ArgGluGluAlaValLeuPheLeu-----ProLys-----LeuAsp 489
QY 1699 CTTGTGGTGGCTGCTGCCCTCCCAAGAGGATGATGAGCTGACCTCTTCTGCTGAAG 1758
    |||:||||| |||: ||| ||| ||| ||| |||
Db 490 Leu-----ProLys-----GlyGluGluValThrIleLeuAlaGln 501
QY 1759 TCCAAAGGGGAACCAATTCACGCTTAGGAACAGTACGCTCCGCCCCACCTCGTG 1818
    |||:||||| |||: ||| ||| ||| ||| |||
Db 502 LysLysLys-----AspValTyArgArgIleVal 511
QY 1819 AACACAAGCTCGGACGACCTTGAG---AGAGGCCATGACACACACAGATGGCAT 1875
    |||:||||| |||: ||| ||| ||| ||| |||
Db 512 GluSerAspValGlyAspSerPheTyrlleArgThrHisPheGluTyGluLysGluSer 531
QY 1876 CTTGGGACCTGAATCATATCACCAGGATCTCAACTCCCTTTGGCCCTGACACAGGCG 1935
    |||:||||| |||: ||| ||| ||| ||| |||
Db 532 ProTyArgLys-----LeuSerPheAsnLysGly 540
QY 1936 CAGATAAGGAACAGCTCGGGCCACTTTTGAAGGCAATGTGGAGGAAGGAGGACAGC 1995
    |||:||||| |||: ||| ||| ||| ||| |||
Db 541 GluValPheArgValValAspThrLeuTyArgAsnGlyLysLeu---GlySerTrpLeuAla 559
QY 1996 AGCCGTTTGGGAGAGATCTCAAGATCAGCTCTCAATCTTCCTTCTTCTCTCTCTCT 2043
    |||:||||| |||: ||| ||| ||| ||| |||
Db 560 IleArgIleGlyLysAsnHisLysGluValGluArgGlyIleIleProAsnLysAsnArg 579
QY 2044 -----CTGGCCAGTGAATTGGTCTCTCCCA-----GCTTGGGGGAC 2082
    |||:||||| |||: ||| ||| ||| ||| |||
Db 580 AlaGluGlnLeuAlaSerValGlnTyThrLeuProLysThrAlaGlyGlyAsp 597

RESULT 7
DLG4_MOUSE
ID DLG4_MOUSE STANDARD; PRT; 724 AA.
AC Q62108;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)
DE (Synapse-associated protein 90) (Discs, large homolog 4).
GN DLG4 OR DLG4H OR PSD95.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DBA/2; TISSUE=Brain;
RA Kohmura N., Yagi T.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
CC -!- SUBUNIT: Interacts with DLGAP1/GKAP (By similarity). Is part of a
CC complex with DLGAP1/GKAP, SHANK1, SHANK2 or SHANK3 (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Concentrated at synaptic
CC junctions primarily on the presynaptic side. Also found in
CC postsynaptic density of neuronal cells (By similarity).
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.

```

```

CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D50621; BAA09297.1; -
DR HSP; P31016; IBE9.
DR MGD; MGI:1277959; Dlg4h.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00072; GuKc; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 65 151 PDZ 1.
FT DOMAIN 160 246 PDZ 2.
FT DOMAIN 313 393 PDZ 3.
FT DOMAIN 428 498 SH3.
FT DOMAIN 534 724 GUANYLATE_KINASE.
SQ SEQUENCE 724 AA; 80472 MW; 7EFFC99E1FF90BA CRC64;

Alignment Scores:
Pred. NO.: 6.5e-07 Length: 724
Score: 258.50 Matches: 127
Percent Similarity: 39.74% Conservative: 90
Best Local Similarity: 23.26% Mismatches: 200
Query Match: 6.27% Indels: 129
DB: 1 Gaps: 25

US-09-502-945-4 (1-2236) x DLG4_MOUSE (1-724)
QY 340 TCCAGGAGCTGAAGGAGTGGTCTGACCGCTCTGCACCCGAGGCGCTCGGCTGAGT 399
    |||:||||| |||: ||| ||| ||| ||| ||| |||
Db 155 AlaGluLysIleIleGluIleLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 173
    |||:||||| |||: ||| ||| ||| ||| ||| |||
QY 400 GTGGCTGGTGGC-----CTGGAGTTTGGCTGTGGGCTCTTTCATCTCCAC 444
    |||:||||| |||: ||| ||| ||| ||| ||| |||
Db 174 IleAlaGlyValGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyValThrLys 193
    |||:||||| |||: ||| ||| ||| ||| ||| |||
QY 445 CTCTCAAAAGCGGTGACGAGACAGCGCTCGG---CTCCAGGTAGGGAGGAGATCGTC 501
    |||:||||| |||: ||| ||| ||| ||| ||| |||
Db 194 IleIleGluGlyGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 213
    |||:||||| |||: ||| ||| ||| ||| ||| |||
QY 502 CGGATCATGATATTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 561
    |||:||||| |||: ||| ||| ||| ||| ||| |||
Db 214 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaLeuLys 233
    |||:||||| |||: ||| ||| ||| ||| ||| |||
QY 562 ---ACCAAGAAACTGTGTCTCATCAAGTG-----AGACACATCGGCTGATCCCC 609
    |||:||||| |||: ||| ||| ||| ||| ||| |||
Db 234 AsnThrTyAspValValTyLeuLysValAlaLysProSerAsnAlaTyLeuSerAsp 253
    |||:||||| |||: ||| ||| ||| ||| ||| |||
QY 610 GTGAAAGCTCTCTGATGAGCCCTCACTTGG---CAGTATGTGATCAGTTGTGTGTCG 666
    |||:||||| |||: ||| ||| ||| ||| ||| |||
Db 254 SerTyAlaProProAspIleThrThrSerTySerGlnHisLeuAspAsnGluIleSer 273
    |||:||||| |||: ||| ||| ||| ||| ||| |||
QY 667 GAATCTGGGGCGCTGCGGAGGAGCGCTGGG----- 696
    |||:||||| |||: ||| ||| ||| ||| ||| |||

```



```

Db 274 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 289
Qy 697 -----TCCCTCGAAATCGGAA-----AACAAAG 720
Db 290 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluGluAspIleProArg 309
Qy 721 GAGAAAGAGGTTCATCAGCCTGGTAGGCTCCGAGCGCTGGTGGTGCAGCATTTCCAGC 780
Db 310 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 329
Qy 781 GGGCCCATCCAGAACGCTGTCATATCATGCCATG-----AAACCTGGCTCC 831
Db 330 GlyGlu---AspGlyGluGlyIlePheHisSerPheIleLeuAlaGlyGlyProAlaAsp 348
Qy 832 CTGCTCTGCTGAGTGGGATGAGATAGGCGACCATGTCGAAGTCAATGGCTGCAC 891
Db 349 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 366
Qy 892 TTCCTAACCTGATCACAAGGAGGCTGTAATGTGCTGAATAATAGCCGAGCTGACC 951
Db 367 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 385
Qy 952 ATCTCCATTGTAGTGCAGCTGGCGGAGCTGTCATGACAGACCGGAGCGCTGGCA 1011
Db 386 ValThrIleIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
Qy 1012 GAGCGCGGCGAGCTGAGCTCCAGCGGAGGAGTTCATGCGAAGACGGCTG----- 1065
Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
Qy 1066 -----GGGANGGAGTCCAAC----- 1080
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
Qy 1080 ----- 1080
Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheHisPheGlyAsp 458
Qy 1081 -----AGATCTCCAGGACGACGAGATGGAGCGGCGGAAAGGAGAAAGAAATTGCC 1134
Db 459 ValLeuHisValIleAspAlaSerAspGluGluTyrTrpGlnAlaArgValHisSer 478
Qy 1135 CAGAAGCGCAGCAGAGGAAATGACAGATACCGGAAGAGATGGAAACAGATGTGACAGG 1194
Db 479 AspSerGluThrAsp-----AspIleGlyPheIleProSerLys 491
Qy 1195 GAAGAGAAGTTTAAGAACCAATGGAA-----GAAGACTGGGGCTCA----- 1236
Db 492 ArgArgValGluArgArgGluTyrSerArgLeuLysAlaLysAspTrpGlySerSer 511
Qy 1237 -----AAGGACACCTACTCTTGCTTAAACCATCACT---GCTGAGGTA 1278
Db 512 GlySerGlnGlyArgGluAspSerValLeuSerTyrGluThrValThrGlnMetGluVal 531
Qy 1279 CAC-----CCAGTACCCTCCGCAAGCAAGATGATCATCAGGAGTGGAACTGAG 1329
Db 532 HisTyrAlaArgProIleIleIleGlyProThrLysAspArgAlaAsnAspAspLeu 551
Qy 1330 CTCGAGCCCGCAGATGACCTTGATGGAGCGCAGGAGGAGGAGGAGGATTTCCGG 1389
Db 552 LeuSerGluPheProAspLysPheSerCysValProHisThrThrArgProLysArg 571
Qy 1390 AAATATGAGGAGGCTTTGACCCCTACTCTATGTTCCACCCAGCAGCAGATCATCGGA 1449
Db 572 GluTyrGluIleAspGlyArgAspTyrHisPheValSerSerArgGluLysMetGluLys 591
Qy 1450 GATCTCGCGCTCTCCATCAGCAGGAGGATCCTTAGACCTGGCCCTGGAGGCGGT 1509
Db 592 AspIleGlnAlaHisLysPheIleGluAlaGlyGlnTyrAsnSerHisLeuTyrGlyThr 611
Qy 1510 GTGACTCCCCCAT-----GGGAAG-----GTGCTC 1536
Db 612 SerValGlnSerValArgGluValAlaGluGlnGlyLysHisCysIleLeuAspValSer 631

```

```

Qy 1537 GTTCTCTGTGTATGAGCGGGAGCTGCTGAGCGCATGCTGCATTGTG----- 1587
Db 632 AlaAsnAlaValArgArgLeuGlnAlaAlaHisLeuHisProIleAlaIlePheIleArg 651
Qy 1588 -----AAAGGGACGAGATGATGGCAATCAACGGCAAGATGTGACAGACTACACCTG 1641
Db 652 ProArgSerLeuGluAsnValLeuGluIleAsnLysArgIleThrGluGluGlnAlaArg 671
Qy 1642 GCTGAGCTGACCTGCC 1659
Db 672 LysAlaPheAspArgAla 677

```

RESULT 8

```

ID Z03_CANFA STANDARD; PRT; 898 AA.
AC 062683;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
DE occludens 3 protein) (Tight junction protein 3).
GN TP3 OR ZO3
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=98198478; PubMed=9531559;
RA Haskins J., Gu L., Wittchen E.S., Hibbard J., Stevenson B.R.;
RT "ZO-3, a novel member of the MAGUK protein family found at the tight
RT junction, interacts with ZO-1 and occludin.";
RL J. Cell Biol. 141:199-208(1998).
CC -1- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: AF023617; AAC39177.1; ..
CC HSSP: P31016; 1BFE.
CC InterPro: IPR000619; Guanylate_kin.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001452; SH3.
CC Pfam: PF00595; PDZ; 3.
CC Pfam: PF00825; Guanylate_kin; 1.
CC SMART: SM00072; GUKC; 1.
CC SMART: SM00228; PDZ; 3.
CC SMART: SM00326; SH3; 1.
CC PROSITE: PS00856; GUANYLATE_KINASE_1; FALSE_NEG.
CC PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
CC PROSITE: PS50106; PDZ; 3.
CC PROSITE: PS50002; SH3; 1.
CC Tight junction; SH3 domain; Repeat; Membrane.
KW DOMAIN 11 93
FT DOMAIN 187 264 PDZ 1.
FT DOMAIN 369 435 PDZ 2.
FT DOMAIN 464 538 PDZ 3.
FT DOMAIN 650 750 GUANYLATE KINASE.
SQ SEQUENCE 898 AA; 98414 MW; 8091D6132DB9F15D CRC64;

```

Alignment Scores:

```

Pred. No.: 2,07e-06 Length: 898
Score: 249.50 Matches: 149

```


Percent Similarity: 37.12% Conservative: 83
 Best Local Similarity: 23.84% Mismatches: 207
 Query Match: 6.06% Indels: 186
 DB: 1 Gaps: 31

US-09-502-945-4 (1-2236) x 203_CANFA (1-898)

```

Qy 385 GGCCTCGGCTGAGTGGTGGTGGCTGGAGTTGGCTGTGGG---CTTCTCATCTCC 441
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 GlyPheGlyIleAlaIleSerGlyArgAspArgAlaSerGlySerValValSer 40
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 442 CACCTCATCAAGCGGTCAGCACACAGCTCGGCTCCAGGTAGGAGGACGATCTGC 501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 AspValValProGlyGlyProAlaAspGly---ArgLeuGlnThrGlyAspHisVal 59
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 502 CGATCAATGATATCTCATCTCCCTGTACCCATGAGGAGTTCATCAACTCATTCGA 561
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 MetValAsnGlyValSerMetGluSerValThrSerThrPheAlaIleGlnLeuLys 79
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 562 ACC---AAGAAACTGTCTCATCAAAAGTG-----AGACATCGGCTGTATCCCC 609
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 ThrCysThrLysLeuAlaAsnIleThrValLysArgProArgLysIleGlnLeuProAla 99
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 610 CTGAAAGCTCT-----CCTGATCAGCCCTCATCTGG 642
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 ThrLysAlaGlyThrSerGlyArgGlyArgGlnGlyLeuGluGluGluAlaAspCysGly 119
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 643 CAG---TATGTGGATCAGTTTGTGCGAATCTGGG----- 675
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 GlnGlyTyrAspGlyAspThrSerSerGlySerGlyArgSerTrpAspLysArgSerArg 139
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 676 -----GGCGTGGAGGCGCTGGCTCCCTCGGAAATCGGAA----- 714
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140 ArgAlaArgThrGlyArgArgAsnGlnAlaGlySerArgGlyArgArgSerProGlyGly 159
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 715 ACACAGGAG----- 723
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 AsnSerGluAlaAsnGlyLeuAlaLeuValSerGlyPheLysArgLeuProArgGlnAsp 179
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 724 -----AAGAGGTCTTCATCAGCCTGGTAGGCTCCGAGGCGCTGGC 765
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 ValHisMetArgProValLysSerValLeuValArgArgThrGluSerGluGluPheGly 199
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 766 TCAGCATTTCCAGCGGCGCCATCCAGAGCCTGTCATCTTTATCAGCCATGTGAACCT 825
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 ValThrLeuGlySer-----GlnIlePheIleLysHisIleThrAsp 213
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 826 GCCTCCCTCTCTGCT---GAGTGGGATGGAGATAGGGACCATGTCTCAAGTCAAT 882
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 SerGlyLeuAlaAlaArgAsnArgGlyLeuGlnGlyAspLeuIleLeuGlnIleAsn 233
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 883 GCGCTGCTCTCTCAACCTGGATCACAGGAGGCTGTAATGTCTGTAATAATAGCCGC 942
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 GlyValSerSerGluAsnLeuSerLeuSerAspThrArgArgLeuIleGluLysSerGlu 253
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 943 ----AGCCTGACCATCTCCATTGTAGCTGCGCTGGCCGGAGCTCTTCATCAGACCGG 999
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 GlyLysLeuThrLeuLeuValLeuArgAspArgGlyGlnPheLeuVal--AsnIleProp 273
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1000 GAGCGGCTGGCAGCGCGCGAGCTGAGCT-----GCAGCGGAGGAG 1044
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 roAlaValSerAspSerAspSerAspSerPheLeuAspAspIleSerAlaLeuGlyS 293
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1045 CTCTCATCGAAGCGGCTGGCGATGGAGTCCCAACAGATCTCCAGGAGCAGGAG 1104
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 erGluLeuSerGlnAlaValProSerHisValProProProProHisAla----- 310
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1105 ATGGAGCGGCAAGGAGAAAGAAATTTGCCAGAGGCGACGAGAGAAATGAGATAC 1164
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 -----GlnArgSerLeu----- 314
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1165 CGGAGGAGATGCAACAGATTCT---AGAGGAGGAGAGAGAGTTTAAGAGCAATGGAA 1221
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 315 --AspSerAspGlyThrAspSerProArgAspSerProProLeu-----A 329
Qy 1222 GAGACTGGGCTCAAAAGAACAGCTACTCTTCCTTAAACCATCACTGCTGAGGTAC 1281
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 rArgGluAsnSerLeuAspSerArgThrIleSer-----GluP 342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1282 CCAGTACCCCTTCGCAAGCCAAAGTATGATCAGGAGTGGAACTGAGCTCGAGCCGCA 1341
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 342 roAspAlaPro-ArgHisSerSerTyrAspIleTyrArgValProSerSerGlnSerAla 361
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1342 GATCAGCTGGATGGAGGACGCGGAGGAGCAGGAGCAGGAGTTCGCGAAATAGAGAA 1401
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 GluAsp-----Arg 364
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1402 GGCTTTGACCCCTACTCTATGTTTACCCACAGCAGATCATGGGAAGGATGTCGGGCTC 1461
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 GlyTyrSerPro-----AspSerArgVal 372
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1462 CTAGCATCAAGAAGAGGAGGATCCTTAGACCTGGCCCTGGAAGCGGTGTGGACTCCCC 1521
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 ValArgPheHisLysGlyThrThrIleGlyLeuArgLeuAlaGlyGlyAsnAsp----- 390
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1522 ATTGGGAAGGTGGTCTGCTGTCTGTATGAGCGGGAGCTGCTGAGCGGCATGTGGC 1581
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 ValGly---IlePheValSerGlyValGlnGlySerProAlaAspGlyGln---Gly 408
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1582 ATTGTGAAGGGGACGAGATCATGCGCAATCAACGGCAAGATTGTGACAGACTACACCTG 1641
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 IleGlnGluGlyAspGlnIleLeuGlnValAsnAspValProPheArgAsnLeuThrArg 428
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1642 GCTGAGCTGACGCTGCCCTGTCAGAACGCTGGAATCAGGGCGGGGACTGGATCGACTT 1701
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 GluGluAla-----ValGlnPhe 434
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1702 GTGTTGTCGCTGCTCCGCCCAAGAGGATGATGACGATGAGCTGACCTTCTTCTCAAGTCC 1761
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 LeuValAlaLeuProProGlyGluGluValGlu-----LeuValThr 448
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1762 AAAAGGGGAAACCAATTCACGCTTAGGAAACAGTGAAGCTCCGCGCCCACTCGTGAAC 1821
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 449 GlnArg-----AsnGluAspIlePheArgLysMetValGln 460
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1822 ACAAGCCTCGGACCGCTTGAGAGAGGCGCATACACACACACAGATGCGATCCTTGG 1881
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 SerArgVal-----GlyAspSerPheTyrIleArgThrHis---Phe 473
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1882 GACCTGAATCTATCACCCAGGAATCTCAAACTCCCTTTGGCCCTGAACGAGGCGAGATA 1941
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 GluLeuGluAlaSerPro-----ProSerGlyLeu-----GlyPheThr 486
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1942 AGGAACAGCTCGGCGCACTTTTGAAGGCCAATGTGGAGGAAAGGGA----- 1989
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 ArgGlyAspValPheHisValLeuAspThrLeuCysProGlyProGlyProSerGlyAla 506
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1990 -----GCAGCCAGCGGTTGGGAGAAGATCTCAAGGATCCAGACTCTCAT 2034
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 ArgGlyThrHisTrpLeuAlaValArgMetGlyArgAspLeuArgGluGlnGluArgGly 526
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2035 TCCTTCTCT 2043
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 527 IleIlePro 529
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 9

DLG1_HUMAN

ID DLG1_HUMAN STANDARD; PRT; 904 AA.

AC Q12959; 012958;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Presynaptic protein SAP97 (Synapse-associated protein 97) (Discs,

DE large homolog 1).

GN DLG1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95024052; PubMed=7937897;
 RA Lue R.A., Marfatia S.M., Branton D., Chishti A.H.;
 RT "Cloning and characterization of hdlg: the human homologue of the
 RT Drosophila discs large tumor suppressor binds to protein 4.1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 450-555.
 RX MEDLINE=96338231; PubMed=87571139;
 RA Cabral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F.,
 RA Marfatia S.M., Chishti A.H., Liddington R.C.;
 RT "Crystal structure of a PDZ domain.";
 RL Nature 382:649-652(1996).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. ASSOCIATES WITH PROTEIN 4.1.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U13897; AAA50599.1; -;
 DR EMBL; U13896; AAA50598.1; -;
 DR PDB; 1PDR; 23-JUL-97.
 DR Genew; HGNC:2900; DLG1.
 DR MIM; 601014;
 DR InterPro; IPR000619; Guanylate_kin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00018; SH3; 1.
 DR Pfam; PF00595; PDZ; 3.
 DR Pfam; PF00625; Guanylate_kin; 1.
 DR Prodom; PD000066; SH3; 1.
 DR SMART; SM00072; GuK; 1.
 DR SMART; SM00228; PDZ; 3.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS50106; PDZ; 3.
 DR PROSITE; PS50002; SH3; 1.
 DR SH3 domain; Repeat; Alternative splicing; 3D-structure.
 DR PDZ 1.
 DR PDZ 2.
 DR PDZ 3.
 DR PDZ 4.
 DR PDZ 5.
 DR PDZ 6.
 DR PDZ 7.
 DR PDZ 8.
 DR PDZ 9.
 DR PDZ 10.
 DR PDZ 11.
 DR PDZ 12.
 DR PDZ 13.
 DR PDZ 14.
 DR PDZ 15.
 DR PDZ 16.
 DR PDZ 17.
 DR PDZ 18.
 DR PDZ 19.
 DR PDZ 20.
 DR PDZ 21.
 DR PDZ 22.
 DR PDZ 23.
 DR PDZ 24.
 DR PDZ 25.
 DR PDZ 26.
 DR PDZ 27.
 DR PDZ 28.
 DR PDZ 29.
 DR PDZ 30.
 DR PDZ 31.
 DR PDZ 32.
 DR PDZ 33.
 DR PDZ 34.
 DR PDZ 35.
 DR PDZ 36.
 DR PDZ 37.
 DR PDZ 38.
 DR PDZ 39.
 DR PDZ 40.
 DR PDZ 41.
 DR PDZ 42.
 DR PDZ 43.
 DR PDZ 44.
 DR PDZ 45.
 DR PDZ 46.
 DR PDZ 47.
 DR PDZ 48.
 DR PDZ 49.
 DR PDZ 50.
 DR PDZ 51.
 DR PDZ 52.
 DR PDZ 53.
 DR PDZ 54.
 DR PDZ 55.
 DR PDZ 56.
 DR PDZ 57.
 DR PDZ 58.
 DR PDZ 59.
 DR PDZ 60.
 DR PDZ 61.
 DR PDZ 62.
 DR PDZ 63.
 DR PDZ 64.
 DR PDZ 65.
 DR PDZ 66.
 DR PDZ 67.
 DR PDZ 68.
 DR PDZ 69.
 DR PDZ 70.
 DR PDZ 71.
 DR PDZ 72.
 DR PDZ 73.
 DR PDZ 74.
 DR PDZ 75.
 DR PDZ 76.
 DR PDZ 77.
 DR PDZ 78.
 DR PDZ 79.
 DR PDZ 80.
 DR PDZ 81.
 DR PDZ 82.
 DR PDZ 83.
 DR PDZ 84.
 DR PDZ 85.
 DR PDZ 86.
 DR PDZ 87.
 DR PDZ 88.
 DR PDZ 89.
 DR PDZ 90.
 DR PDZ 91.
 DR PDZ 92.
 DR PDZ 93.
 DR PDZ 94.
 DR PDZ 95.
 DR PDZ 96.
 DR PDZ 97.
 DR PDZ 98.
 DR PDZ 99.
 DR PDZ 100.
 DR PDZ 101.
 DR PDZ 102.
 DR PDZ 103.
 DR PDZ 104.
 DR PDZ 105.
 DR PDZ 106.
 DR PDZ 107.
 DR PDZ 108.
 DR PDZ 109.
 DR PDZ 110.
 DR PDZ 111.
 DR PDZ 112.
 DR PDZ 113.
 DR PDZ 114.
 DR PDZ 115.
 DR PDZ 116.
 DR PDZ 117.
 DR PDZ 118.
 DR PDZ 119.
 DR PDZ 120.
 DR PDZ 121.
 DR PDZ 122.
 DR PDZ 123.
 DR PDZ 124.
 DR PDZ 125.
 DR PDZ 126.
 DR PDZ 127.
 DR PDZ 128.
 DR PDZ 129.
 DR PDZ 130.
 DR PDZ 131.
 DR PDZ 132.
 DR PDZ 133.
 DR PDZ 134.
 DR PDZ 135.
 DR PDZ 136.
 DR PDZ 137.
 DR PDZ 138.
 DR PDZ 139.
 DR PDZ 140.
 DR PDZ 141.
 DR PDZ 142.
 DR PDZ 143.
 DR PDZ 144.
 DR PDZ 145.
 DR PDZ 146.
 DR PDZ 147.
 DR PDZ 148.
 DR PDZ 149.
 DR PDZ 150.
 DR PDZ 151.
 DR PDZ 152.
 DR PDZ 153.
 DR PDZ 154.
 DR PDZ 155.
 DR PDZ 156.
 DR PDZ 157.
 DR PDZ 158.
 DR PDZ 159.
 DR PDZ 160.
 DR PDZ 161.
 DR PDZ 162.
 DR PDZ 163.
 DR PDZ 164.
 DR PDZ 165.
 DR PDZ 166.
 DR PDZ 167.
 DR PDZ 168.
 DR PDZ 169.
 DR PDZ 170.
 DR PDZ 171.
 DR PDZ 172.
 DR PDZ 173.
 DR PDZ 174.
 DR PDZ 175.
 DR PDZ 176.
 DR PDZ 177.
 DR PDZ 178.
 DR PDZ 179.
 DR PDZ 180.
 DR PDZ 181.
 DR PDZ 182.
 DR PDZ 183.
 DR PDZ 184.
 DR PDZ 185.
 DR PDZ 186.
 DR PDZ 187.
 DR PDZ 188.
 DR PDZ 189.
 DR PDZ 190.
 DR PDZ 191.
 DR PDZ 192.
 DR PDZ 193.
 DR PDZ 194.
 DR PDZ 195.
 DR PDZ 196.
 DR PDZ 197.
 DR PDZ 198.
 DR PDZ 199.
 DR PDZ 200.
 DR PDZ 201.
 DR PDZ 202.
 DR PDZ 203.
 DR PDZ 204.
 DR PDZ 205.
 DR PDZ 206.
 DR PDZ 207.
 DR PDZ 208.
 DR PDZ 209.
 DR PDZ 210.
 DR PDZ 211.
 DR PDZ 212.
 DR PDZ 213.
 DR PDZ 214.
 DR PDZ 215.
 DR PDZ 216.
 DR PDZ 217.
 DR PDZ 218.
 DR PDZ 219.
 DR PDZ 220.
 DR PDZ 221.
 DR PDZ 222.
 DR PDZ 223.
 DR PDZ 224.
 DR PDZ 225.
 DR PDZ 226.
 DR PDZ 227.
 DR PDZ 228.
 DR PDZ 229.
 DR PDZ 230.
 DR PDZ 231.
 DR PDZ 232.
 DR PDZ 233.
 DR PDZ 234.
 DR PDZ 235.
 DR PDZ 236.
 DR PDZ 237.
 DR PDZ 238.
 DR PDZ 239.
 DR PDZ 240.
 DR PDZ 241.
 DR PDZ 242.
 DR PDZ 243.
 DR PDZ 244.
 DR PDZ 245.
 DR PDZ 246.
 DR PDZ 247.
 DR PDZ 248.
 DR PDZ 249.
 DR PDZ 250.
 DR PDZ 251.
 DR PDZ 252.
 DR PDZ 253.
 DR PDZ 254.
 DR PDZ 255.
 DR PDZ 256.
 DR PDZ 257.
 DR PDZ 258.
 DR PDZ 259.
 DR PDZ 260.
 DR PDZ 261.
 DR PDZ 262.
 DR PDZ 263.
 DR PDZ 264.
 DR PDZ 265.
 DR PDZ 266.
 DR PDZ 267.
 DR PDZ 268.
 DR PDZ 269.
 DR PDZ 270.
 DR PDZ 271.
 DR PDZ 272.
 DR PDZ 273.
 DR PDZ 274.
 DR PDZ 275.
 DR PDZ 276.
 DR PDZ 277.
 DR PDZ 278.
 DR PDZ 279.
 DR PDZ 280.
 DR PDZ 281.
 DR PDZ 282.
 DR PDZ 283.
 DR PDZ 284.
 DR PDZ 285.
 DR PDZ 286.
 DR PDZ 287.
 DR PDZ 288.
 DR PDZ 289.
 DR PDZ 290.
 DR PDZ 291.
 DR PDZ 292.
 DR PDZ 293.
 DR PDZ 294.
 DR PDZ 295.
 DR PDZ 296.
 DR PDZ 297.
 DR PDZ 298.
 DR PDZ 299.
 DR PDZ 300.
 DR PDZ 301.
 DR PDZ 302.
 DR PDZ 303.
 DR PDZ 304.
 DR PDZ 305.
 DR PDZ 306.
 DR PDZ 307.
 DR PDZ 308.
 DR PDZ 309.
 DR PDZ 310.
 DR PDZ 311.
 DR PDZ 312.
 DR PDZ 313.
 DR PDZ 314.
 DR PDZ 315.
 DR PDZ 316.
 DR PDZ 317.
 DR PDZ 318.
 DR PDZ 319.
 DR PDZ 320.
 DR PDZ 321.
 DR PDZ 322.
 DR PDZ 323.
 DR PDZ 324.
 DR PDZ 325.
 DR PDZ 326.
 DR PDZ 327.
 DR PDZ 328.
 DR PDZ 329.
 DR PDZ 330.
 DR PDZ 331.
 DR PDZ 332.
 DR PDZ 333.
 DR PDZ 334.
 DR PDZ 335.
 DR PDZ 336.
 DR PDZ 337.
 DR PDZ 338.
 DR PDZ 339.
 DR PDZ 340.
 DR PDZ 341.
 DR PDZ 342.
 DR PDZ 343.
 DR PDZ 344.
 DR PDZ 345.
 DR PDZ 346.
 DR PDZ 347.
 DR PDZ 348.
 DR PDZ 349.
 DR PDZ 350.
 DR PDZ 351.
 DR PDZ 352.
 DR PDZ 353.
 DR PDZ 354.
 DR PDZ 355.
 DR PDZ 356.
 DR PDZ 357.
 DR PDZ 358.
 DR PDZ 359.
 DR PDZ 360.
 DR PDZ 361.
 DR PDZ 362.
 DR PDZ 363.
 DR PDZ 364.
 DR PDZ 365.
 DR PDZ 366.
 DR PDZ 367.
 DR PDZ 368.
 DR PDZ 369.
 DR PDZ 370.
 DR PDZ 371.
 DR PDZ 372.
 DR PDZ 373.
 DR PDZ 374.
 DR PDZ 375.
 DR PDZ 376.
 DR PDZ 377.
 DR PDZ 378.
 DR PDZ 379.
 DR PDZ 380.
 DR PDZ 381.
 DR PDZ 382.
 DR PDZ 383.
 DR PDZ 384.
 DR PDZ 385.
 DR PDZ 386.
 DR PDZ 387.
 DR PDZ 388.
 DR PDZ 389.
 DR PDZ 390.
 DR PDZ 391.
 DR PDZ 392.
 DR PDZ 393.
 DR PDZ 394.
 DR PDZ 395.
 DR PDZ 396.
 DR PDZ 397.
 DR PDZ 398.
 DR PDZ 399.
 DR PDZ 400.
 DR PDZ 401.
 DR PDZ 402.
 DR PDZ 403.
 DR PDZ 404.
 DR PDZ 405.
 DR PDZ 406.
 DR PDZ 407.
 DR PDZ 408.
 DR PDZ 409.
 DR PDZ 410.
 DR PDZ 411.
 DR PDZ 412.
 DR PDZ 413.
 DR PDZ 414.
 DR PDZ 415.
 DR PDZ 416.
 DR PDZ 417.
 DR PDZ 418.
 DR PDZ 419.
 DR PDZ 420.
 DR PDZ 421.
 DR PDZ 422.
 DR PDZ 423.
 DR PDZ 424.
 DR PDZ 425.
 DR PDZ 426.
 DR PDZ 427.
 DR PDZ 428.
 DR PDZ 429.
 DR PDZ 430.
 DR PDZ 431.
 DR PDZ 432.
 DR PDZ 433.
 DR PDZ 434.
 DR PDZ 435.
 DR PDZ 436.
 DR PDZ 437.
 DR PDZ 438.
 DR PDZ 439.
 DR PDZ 440.
 DR PDZ 441.
 DR PDZ 442.
 DR PDZ 443.
 DR PDZ 444.
 DR PDZ 445.
 DR PDZ 446.
 DR PDZ 447.
 DR PDZ 448.
 DR PDZ 449.
 DR PDZ 450.
 DR PDZ 451.
 DR PDZ 452.
 DR PDZ 453.
 DR PDZ 454.
 DR PDZ 455.
 DR PDZ 456.
 DR PDZ 457.
 DR PDZ 458.
 DR PDZ 459.
 DR PDZ 460.
 DR PDZ 461.
 DR PDZ 462.
 DR PDZ 463.
 DR PDZ 464.
 DR PDZ 465.
 DR PDZ 466.
 DR PDZ 467.
 DR PDZ 468.
 DR PDZ 469.
 DR PDZ 470.
 DR PDZ 471.
 DR PDZ 472.
 DR PDZ 473.
 DR PDZ 474.
 DR PDZ 475.
 DR PDZ 476.
 DR PDZ 477.
 DR PDZ 478.
 DR PDZ 479.
 DR PDZ 480.
 DR PDZ 481.
 DR PDZ 482.
 DR PDZ 483.
 DR PDZ 484.
 DR PDZ 485.
 DR PDZ 486.
 DR PDZ 487.
 DR PDZ 488.
 DR PDZ 489.
 DR PDZ 490.
 DR PDZ 491.
 DR PDZ 492.
 DR PDZ 493.
 DR PDZ 494.
 DR PDZ 495.
 DR PDZ 496.
 DR PDZ 497.
 DR PDZ 498.
 DR PDZ 499.
 DR PDZ 500.
 DR PDZ 501.
 DR PDZ 502.
 DR PDZ 503.
 DR PDZ 504.
 DR PDZ 505.
 DR PDZ 506.
 DR PDZ 507.
 DR PDZ 508.
 DR PDZ 509.
 DR PDZ 510.
 DR PDZ 511.
 DR PDZ 512.
 DR PDZ 513.
 DR PDZ 514.
 DR PDZ 515.
 DR PDZ 516.
 DR PDZ 517.
 DR PDZ 518.
 DR PDZ 519.
 DR PDZ 520.
 DR PDZ 521.
 DR PDZ 522.
 DR PDZ 523.
 DR PDZ 524.
 DR PDZ 525.
 DR PDZ 526.
 DR PDZ 527.
 DR PDZ 528.
 DR PDZ 529.
 DR PDZ 530.
 DR PDZ 531.
 DR PDZ 532.
 DR PDZ 533.
 DR PDZ 534.
 DR PDZ 535.
 DR PDZ 536.
 DR PDZ 537.
 DR PDZ 538.
 DR PDZ 539.
 DR PDZ 540.
 DR PDZ 541.
 DR PDZ 542.
 DR PDZ 543.
 DR PDZ 544.
 DR PDZ 545.
 DR PDZ 546.
 DR PDZ 547.
 DR PDZ 548.
 DR PDZ 549.
 DR PDZ 550.
 DR PDZ 551.
 DR PDZ 552.
 DR PDZ 553.
 DR PDZ 554.
 DR PDZ 555.
 DR PDZ 556.
 DR PDZ 557.
 DR PDZ 558.
 DR PDZ 559.
 DR PDZ 560.
 DR PDZ 561.
 DR PDZ 562.
 DR PDZ 563.
 DR PDZ 564.
 DR PDZ 565.
 DR PDZ 566.
 DR PDZ 567.
 DR PDZ 568.
 DR PDZ 569.
 DR PDZ 570.
 DR PDZ 571.
 DR PDZ 572.
 DR PDZ 573.
 DR PDZ 574.
 DR PDZ 575.
 DR PDZ 576.
 DR PDZ 577.
 DR PDZ 578.
 DR PDZ 579.
 DR PDZ 580.
 DR PDZ 581.
 DR PDZ 582.
 DR PDZ 583.
 DR PDZ 584.
 DR PDZ 585.
 DR PDZ 586.
 DR PDZ 587.
 DR PDZ 588.
 DR PDZ 589.
 DR PDZ 590.
 DR PDZ 591.
 DR PDZ 592.
 DR PDZ 593.
 DR PDZ 594.
 DR PDZ 595.
 DR PDZ 596.
 DR PDZ 597.
 DR PDZ 598.
 DR PDZ 599.
 DR PDZ 600.
 DR PDZ 601.
 DR PDZ 602.
 DR PDZ 603.
 DR PDZ 604.
 DR PDZ 605.
 DR PDZ 606.
 DR PDZ 607.
 DR PDZ 608.
 DR PDZ 609.
 DR PDZ 610.
 DR PDZ 611.
 DR PDZ 612.
 DR PDZ 613.
 DR PDZ 614.
 DR PDZ 615.
 DR PDZ 616.
 DR PDZ 617.
 DR PDZ 618.
 DR PDZ 619.
 DR PDZ 620.
 DR PDZ 621.
 DR PDZ 622.
 DR PDZ 623.
 DR PDZ 624.
 DR PDZ 625.
 DR PDZ 626.
 DR PDZ 627.
 DR PDZ 628.
 DR PDZ 629.
 DR PDZ 630.
 DR PDZ 631.
 DR PDZ 632.
 DR PDZ 633.
 DR PDZ 634.
 DR PDZ 635.
 DR PDZ 636.
 DR PDZ 637.
 DR PDZ 638.
 DR PDZ 639.
 DR PDZ 640.
 DR PDZ 641.
 DR PDZ 642.
 DR PDZ 643.
 DR PDZ 644.
 DR PDZ 645.
 DR PDZ 646.
 DR PDZ 647.
 DR PDZ 648.
 DR PDZ 649.
 DR PDZ 650.
 DR PDZ 651.
 DR PDZ 652.
 DR PDZ 653.
 DR PDZ 654.
 DR PDZ 655.
 DR PDZ 656.
 DR PDZ 657.
 DR PDZ 658.
 DR PDZ 659.
 DR PDZ 660.
 DR PDZ 661.
 DR PDZ 662.
 DR PDZ 663.
 DR PDZ 664.
 DR PDZ 665.
 DR PDZ 666.
 DR PDZ 667.
 DR PDZ 668.
 DR PDZ 669.
 DR PDZ 670.
 DR PDZ 671.
 DR PDZ 672.
 DR PDZ 673.
 DR PDZ 674.
 DR PDZ 675.
 DR PDZ 676.
 DR PDZ 677.
 DR PDZ 678.
 DR PDZ 679.
 DR PDZ 680.
 DR PDZ 681.
 DR PDZ 682.
 DR PDZ 683.
 DR PDZ 684.
 DR PDZ 685.
 DR PDZ 686.
 DR PDZ 687.
 DR PDZ 688.
 DR PDZ 689.
 DR PDZ 690.
 DR PDZ 691.
 DR PDZ 692.
 DR PDZ 693.
 DR PDZ 694.
 DR PDZ 695.
 DR PDZ 696.
 DR PDZ 697.
 DR PDZ 698.
 DR PDZ 699.
 DR PDZ 700.
 DR PDZ 701.
 DR PDZ 702.
 DR PDZ 703.
 DR PDZ 704.
 DR PDZ 705.
 DR PDZ 706.
 DR PDZ 707.
 DR PDZ 708.
 DR PDZ 709.
 DR PDZ 710.
 DR PDZ 711.
 DR PDZ 712.
 DR PDZ 713.
 DR PDZ 714.
 DR PDZ 715.
 DR PDZ 716.
 DR PDZ 717.
 DR PDZ 718.
 DR PDZ 719.
 DR PDZ 720.
 DR PDZ 721.
 DR PDZ 722.
 DR PDZ 723.
 DR PDZ 724.
 DR PDZ 725.
 DR PDZ 726.
 DR PDZ 727.
 DR PDZ 728.
 DR PDZ 729.
 DR PDZ 730.
 DR PDZ 731.
 DR PDZ 732.
 DR PDZ 733.
 DR PDZ 734.
 DR PDZ 735.
 DR PDZ 736.
 DR PDZ 737.
 DR PDZ 738.
 DR PDZ 739.
 DR PDZ 740.
 DR PDZ 741.
 DR PDZ 742.
 DR PDZ 743.
 DR PDZ 744.
 DR PDZ 745.
 DR PDZ 746.
 DR PDZ 747.
 DR PDZ 748.
 DR PDZ 749.
 DR PDZ 750.
 DR PDZ 751.
 DR PDZ 752.
 DR PDZ 753.
 DR PDZ 754.
 DR PDZ 755.
 DR PDZ 756.
 DR PDZ 757.
 DR PDZ 758.
 DR PDZ 759.
 DR PDZ 760.
 DR PDZ 761.
 DR PDZ 762.
 DR PDZ 763.
 DR PDZ 764.
 DR PDZ 765.
 DR PDZ 766.
 DR PDZ 767.
 DR PDZ 768.
 DR PDZ 769.
 DR PDZ 770.
 DR PDZ 771.
 DR PDZ 772.
 DR PDZ 773.
 DR PDZ 774.
 DR PDZ 775.
 DR PDZ 776.
 DR PDZ 777.
 DR PDZ 778.
 DR PDZ 779.
 DR PDZ 780.
 DR PDZ 781.
 DR PDZ 782.
 DR PDZ 783.
 DR PDZ 784.
 DR PDZ 785.
 DR PDZ 786.
 DR PDZ 787.
 DR PDZ 788.
 DR PDZ 789.
 DR PDZ 790.
 DR PDZ 791.
 DR PDZ 792.
 DR PDZ 793.
 DR PDZ 794.
 DR PDZ 795.
 DR PDZ 796.
 DR PDZ 797.
 DR PDZ 798.
 DR PDZ 799.
 DR PDZ 800.
 DR PDZ 801.
 DR PDZ 802.
 DR PDZ 803.
 DR PDZ 804.
 DR PDZ 805.
 DR PDZ 806.
 DR PDZ 807.
 DR PDZ 808.
 DR PDZ 809.
 DR PDZ 810.
 DR PDZ 811.
 DR PDZ 812.
 DR PDZ 813.
 DR PDZ 814.
 DR PDZ 815.
 DR PDZ 816.
 DR PDZ 817.
 DR PDZ 818.
 DR PDZ 819.
 DR PDZ 820.
 DR PDZ 821.
 DR PDZ 822.
 DR PDZ 823.
 DR PDZ 824.
 DR PDZ 825.
 DR PDZ 826.
 DR PDZ 827.
 DR PDZ 828.
 DR PDZ 829.
 DR PDZ 830.
 DR PDZ 831.
 DR PDZ 832.
 DR PDZ 833.
 DR PDZ 834.
 DR PDZ 835.
 DR PDZ 836.
 DR PDZ 837.
 DR PDZ 838.
 DR PDZ 839.
 DR PDZ 840.
 DR PDZ 841.
 DR PDZ 842.
 DR PDZ 843.
 DR PDZ 844.
 DR PDZ 845.
 DR PDZ 846.
 DR PDZ 847.
 DR PDZ 848.
 DR PDZ 849.
 DR PDZ 850.
 DR PDZ 851.
 DR PDZ 852.
 DR PDZ 853.
 DR PDZ 854.
 DR PDZ 855.
 DR PDZ 856.
 DR PDZ 857.
 DR PDZ 858.
 DR PDZ 859.
 DR PDZ 860.
 DR PDZ 861.
 DR PDZ 862.
 DR PDZ 863.
 DR PDZ 864.
 DR PDZ 865.
 DR PDZ 866.
 DR PDZ 867.
 DR PDZ 868.
 DR PDZ 869.
 DR PDZ 870.
 DR PDZ 871.
 DR PDZ 872.
 DR PDZ 873.
 DR PDZ 874.
 DR PDZ 875.
 DR PDZ 876.
 DR PDZ 877.
 DR PDZ 878.
 DR PDZ 879.
 DR PDZ 880.
 DR PDZ 881.
 DR PDZ 882.
 DR PDZ 883.
 DR PDZ 884.
 DR PDZ 885.
 DR PDZ 886.
 DR PDZ 887.
 DR PDZ 888.
 DR PDZ 889.
 DR PDZ 890.
 DR PDZ 891.
 DR PDZ 892.
 DR PDZ 893.
 DR PDZ 894.
 DR PDZ 895.
 DR PDZ 896.
 DR PDZ 897.
 DR PDZ 898.
 DR PDZ 899.
 DR PDZ 900.
 DR PDZ 901.
 DR PDZ 902.
 DR PDZ 903.
 DR PDZ 904.
 DR PDZ 905.
 DR PDZ 906.
 DR PDZ 907.
 DR PDZ 908.
 DR PDZ 909.
 DR PDZ 910.
 DR PDZ 911.
 DR PDZ 912.
 DR PDZ 913.
 DR PDZ 914.
 DR PDZ 915.
 DR PDZ 916.
 DR PDZ 917.
 DR PDZ 918.
 DR PDZ 919.
 DR PDZ 920.
 DR PDZ 921.
 DR PDZ 922.
 DR PDZ 923.
 DR PDZ 924.
 DR PDZ 925.
 DR PDZ 926.
 DR PD


```

Db 663 LysThrArgAspLysGlyGluLeuProAspMetGlySerLysGlyLeuLysHisVal 682
Qy 1240 -----
Db 683 ThrSerAsnAlaSerAspSerCluSerSerTyrArgGlyGlnGluGluTyrValLeuSer 702
Qy 1255 CCTAAACCATCATCTGCTGAG-----GTFACACCCAGTACCCTTCGCAAGCCA 1302
Db 703 TyrGluProValAsnGlnGlnValAsnTyrThrArgProValIleIleLeuGlyPro 722
Qy 1303 AAGTATGATCAGGAGTGAACCTGAGCTCGAGCCGCGAGATGACCTGGATGGAGGCACG 1362
Db 723 MetLysAspArgIleAsnAspLeuIleSerGluPheProAspLysPheGlySerCys 742
Qy 1363 GAGGACGAGGAGACGAGATTTCGGAATATGAGGAAGGCTTTCGACCCCTACTCTATG 1422
Db 743 ValProHisThrThrArgProLysArgAspTyrGluValAspGlyArgAspTyrHisPhe 762
Qy 1423 TTCACCCAGCAGCATCATCGGGAAGGATGTCGGCTCCCTACGATCAAGAAGAGGGA 1482
Db 763 ValThrSerArgGluGlnMetGluLysAspIleGlnGluHisLysPheIleGluAlaGly 782
Qy 1483 TCCTTAGACTGGCCCTGGAAGCGGTGTGGACTCCGCCATTTGGGAAGTGGTCTTTCT 1542
Db 783 GlnTyrAsnAsnHisLeuTyrGly-----ThrSerValGln 794
Qy 1543 GCTGTGTATGACGGGGAGCTGCTGAGCGGCATGTGGCATTTGTGAAAGGGACGAGATC 1602
Db 795 SerValArgGluValAlaGlyLysGlyLysHisCys-----Ile 807
Qy 1603 ATGGCAATCAACGCAAGATTGTGACAGACTACACCTGGCTGAG 1647
Db 808 LeuAspValSerGlyAsnAlaIleLysArgLeuGlnIleAlaGln 822

RESULT 10
DLG3_MOUSE
ID DLG3_MOUSE STANDARD; PRT; 849 AA.
AC P01175;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic protein SAPI02 (Synapse-associated protein 102) (Discs,
DE large homolog 3).
GN DLG3 OR DLG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Kohmura N., Makino S., Yagi T.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC SUBUNIT NP2B.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D87117; BAA13249.1; .
DR HSSP; Q12959; 1PDR.
DR MGD; MGI:1888986; Dlg3.
DR InterPro; IPR000619; Guanylate_kin.

```

```

DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 149 235 PDZ 1.
FT DOMAIN 244 330 PDZ 2.
FT DOMAIN 404 484 PDZ 3.
FT DOMAIN 519 589 SH3.
FT DOMAIN 659 849 GUANYLATE_KINASE.
SQ SEQUENCE 849 AA; 93482 MW; EF3EF2D7513538EE CRC64;

Alignment Scores:
Pred. No.: 4,6e-06 Length: 849
Score: 243.00 Matches: 137
Percent Similarity: 36.29% Conservative: 96
Best Local Similarity: 21.34% Mismatches: 191
Query Match: 5.90% Indels: 218
DB: 1 Gaps: 29

US-09-502-945-4 (1-2236) x DLG3_MOUSE (1-849)
Qy 8 CGTTCGGCTCGCGCTCTTT-----CCAGTCTCTGGCAGCCGG 46
Db 70 ArgThrLysAlaLysLeuIleProThrGlyArgAspValGlyProValProLysPro 89
Qy 47 GCACCCGAGAACGGTCTGTCGAACGACGACGCTGACCTGGCCCGCCAGCA----TGGACC 103
Db 90 ValPro-----GlyLysSerThrProLysLeuAsnGlySerGlyProGlyTyr--- 105
Qy 104 GAAAAGTGGCCCGAGAATTCGGCATAAGGTGGATTTTCTGATTGAAATGATGACAGAGA 163
Db 106 -----TrpProGlu----- 108
Qy 164 AGGACTATCTCTATGATGTCTGCGAATGTACACACGACCATGGACGTGGCCGTGCTCG 223
Db 109 -----CysThrAsnArgAspTyr----- 116
Qy 224 TGGGAGACCTGAAGCTGTGTCATCAATGAACCCAGCGTCTGCTCTCTTTGATGCCATTC 283
Db 117 -----TyrGluIleAlaSerProAlaPro---LeuLeuValAsnPro-Gl 130
Qy 284 GCGCGCTGATCCCACTGAAGCACCAGGTGGGAATATGATACAGTACGCCCGCGCTCCA 343
Db 130 uAlaLeuGluProSerLeuSerValAsnGly-SerAspGlyMetPhe----- 145
Qy 344 GGAAGCTGAAGAGGTGGTCTGACCGCTGACCCCGAAGCGCTGGCTGAGTGTGC 403
Db 146 --LysTyrGluGluIleValLeuGluArgGlyAsn---SerGlyLeuGlyPheSerIleA 164
Qy 404 GTGGTGGCTCGAG-----TTTGCTGTGGGCTCTTCATCTCCACCTCA 448
Db 164 laGlyIleAspAsnProHisValProAspProGlyIlePheIleThrLysIle 184
Qy 449 TCAAGGCGGTCAGCAGACAGCGTCGGG---CTCCAGGTAGGGGAGAGATCTCGGA 505
Db 184 leProGlyGlyAlaAlaAlaMetAspGlyArgLeuGlyValAsnAspCysValLeuArgV 204
Qy 506 TCAATGGATATTCATCTCTCTGTCCTGTCACCCATGAGGAGGTTCATCAACCTCATCGAACCA 565
Db 204 alAsnGluValAspValSerGluValHisSerArgAlaValGluAlaLeuLysGluA 224
Qy 566 AGAAACTGTGTCTCCATCAAGTGAGACACATCGCGCTGATCCCGCTGAAAGCTCTCTCTG 625

```



```

Db 224 laGlyProVal-----ValArg-----LeuValValArgArgGlnP 237
Qy 626 ATGAGCCCTCACTGGCAGTATGTCATGTTGTGTCGGAATCTGGGCGCTGCAG 685
Db 237 roProProGluThrIleMetGluValAsn----- 246
Qy 686 GCAGCGCTGGCTCCCTCGAAATCGGAAACAAGGAGAAGAGCTTCATCAGCCCTG 745
Db 247 -----LeuLeuL 249
Qy 746 TAGGCTCCGAGCGCTTGCTGCAGCATTCCTCAGCGGCCCTCAGAG-----CTG 799
Db 249 ysGlyProLysGlyLeuGlyPheSerIleAlaGlyIleGlyAsnGlnHisIleProG 269
Qy 800 GC-----ATCTTATACCCATGTCAACCTGGCTCCCTGCTGCTGAGGTGGA- 849
Db 269 lyAspAsnSerIleThrIleThrLysIleIleGlyAlaAlaGlnLysAspGlyA 289
Qy 850 --TTGAGATAGGACGATGTTCGAAGTCAATGGCTGCAGCTCTCTAACCTGGATC 907
Db 289 rgLeuGlnIleGlyAspArgLeuLeuAlaValAlaAsnThrAsnLeuGlnAspValArgH 309
Qy 908 ACAAGGAGCTTAATGTCTGAATAATAGCCGAGCTGACCATCTCCATGTAGCTG 967
Db 309 IsGluGluAlaValAlaSerLeuLysAsnThrSerAspMetValTyrLeuLysValAlaL 329
Qy 968 CAGCTGGCGGAGCTGTTTCATGACAGACCGGAGCGGTGTCAGAGCGCGGAGCTG 1027
Db 329 ysProGly-----S 332
Qy 1028 AGCTGCAGCGGAGGAGCTTCTCATGCAGAGCGGTGCGGATGGAGTCCAACAAGATCC 1087
Db 332 erIleHisLeuAsnAspMetTyrAlaProProAspTyrAlaSerThrPheThrAlaLeu- 351
Qy 1088 TCCAGGACGAGGAGATGGAGCGGCAAGAGAGAAATGCCAGAGGAGGAGCAG 1147
Db 352 -----AlaA 353
Qy 1148 AGGAAATGAGAGATACCGGAGGAGATGGACAGATTTAGAGGAGGAGAGAGTTTA 1207
Db 353 spAsnHisIleSerHisAsnSerSerLeuGlyTyrLeuGlyAlaValGluSerLysVal- 372
Qy 1208 AGAGCAATGGGAAGAGATGGGGCTCAAGGAACAGCTACTCTTCCTTAAACCATCA 1267
Db 372 ----- 372
Qy 1268 CTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATGATCAGGAGTGGAACTG 1327
Db 373 -----ThrTyrProAlaPro-----ProG 379
Qy 1328 AGCTGAGCGGCAGATGACCTGGATGGAGGACGCGGAGGAGGAGGAGGATTTCC 1387
Db 379 InValProProthr----- 383
Qy 1388 GGAATATGAGGAGGCTTTGACCCCTACTCTATGTTCCACCCAGCAGCAGATCATGGGA 1447
Db 384 --ArgTyrSerPro-----IleProArgHisMetLeuAlaGluGluAspPheThrArg 401
Qy 1448 AGGATGTCGGCTCTCAGCATCAAGAGAGGAGGATCCTTAGACTGCGCTGGAAGCG 1507
Db 401 luProArgLysIleIleLeuHisLysGlySerThrGlyLeuGlyPheAsnIleValGlyG 421
Qy 1508 GTGTGACTCCCTTGGGAGGTGGTCTTCTGCTGCTGATGAGCGGGAGCTGCTG 1567
Db 421 lyGluAspGlyGlu-----GlyIlePheValSerPheIleLeuAlaGlyGlyProAlaA 439
Qy 1568 AGCGCATGGGCGATTCTGAAAGGAGGAGATCATGGCATCAACCGCAGATTTGTA 1627
Db 439 spLeuSerGlyGluLeuArgArgGlyAspArgIleLeuSerValAsnGlyValAsnLeuA 459
Qy 1628 CAGACTACACCTGCTGAGGTGACCTGCGCTGCGCAAGAGCGCTGGAATCAGGCGGG 1687
Db 459 rgAsnAlaThrHisGluGlnAlaAlaAlaLeuLysArgAla-----GlyG 475

```

```

Qy 1688 ACTGATCGACCTTGTTGTTGCTGCGTCCGCCCCCAAGAGGATATGACGATGAGTGACCT 1747
Db 475 InSerVal---ThrIleValAlaGlnTyrArgProGluGluTyr----- 488
Qy 1748 TCTTGCTGAAGTCCAAAGAGGGAACCAAAATTCACGCGCTTAGGAAACAGTACGCTCCGC 1807
Db 489 -----SerArgPheGluSerLysIleHis-----AspLeuArg 500
Qy 1808 CCCACCTCGTGAACACAAAGCCT-----CGGACCAAGCTTGAGA 1846
Db 500 luGlnMetMetAsnSerSerMetSerSerGlySerGlySerLeuArgThrSerGluLysA 520
Qy 1847 GA 1848
Db 520 rg 520
RESULT 11
ID_DLG3_RAT STANDARD; PRT; 849 AA.
AC Q62936; P70547;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic protein SAP102 (Synapse-associated protein 102) (PSD-
DE 95/SAP90 related protein 1) (Discs, large homolog 3).
GN DLG3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE=Brain;
RX MEDLINE=96374358; PubMed=8780649;
RA Mueller B.M., Kistner U., Kindler S., Chung W.J., Kuhlendahl S.,
RA Fenster S.D., Lau L.-F., Veh R.W., Haganir R.L., Gundelfinger E.D.,
RA Garner C.C.;
RT "SAP102, a novel postsynaptic protein that interacts with NMDA
RT receptor complexes in vivo.";
RL Neuron 17:255-265(1996).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RA Irie M., Hata Y., Takai Y.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC SUBUNIT NR2B.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U50147; AAA93031.1; -.
CC EMBL; U53367; AAB48561.1; -.
CC HSP; Q12959; IPDR.
CC InterPro; IPR000619; Guanylate_Kin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00595; PDZ; 3.
CC Pfam; PF00625; Guanylate_kin; 1.
CC ProDom; PD000066; SH3; 1.
CC SMART; SM00072; GUKc; 1.
CC SMART; SM00228; PDZ; 3.
CC SMART; SM00326; SH3; 1.

```


QY 1832 GGACGAGCTTGAGAGA 1848
 Db 515 rgThrSerGluLysArg 520

RESULT 12

DLGL_RAT STANDARD; PRT; 911 AA.
 AC Q62696;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Presynaptic protein SAP97 (Synapse-associated protein 97) (Discs,
 DE large homolog 1).
 GN DLGL
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95198112; PubMed=7891172;
 RA Mueller B.M., Kistner U., Veh R.W., Cases-Langhoff C., Becker B.,
 RA Gundelfinger E.D., Garner C.C.;
 RT "Molecular characterization and spatial distribution of SAP97, a
 RT novel presynaptic protein homologous to SAP90 and the Drosophila
 RT discs-large tumor suppressor protein.";
 RL J. Neurosci. 15:2354-2366(1995).
 CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS. MAY PLAY A ROLE IN
 CC CELL ADHESION
 CC -!- SUBCELLULAR LOCATION: IN CULTURED T84 CELLS, IT IS RESTRICTED TO
 CC THE CYTOPLASMIC SURFACE OF THE PLASMA MEMBRANES BETWEEN ADJACENT
 CC CELLS, BUT NOT AT THE EDGES OF CELLS LACKING CELL-CELL CONTACT.
 CC -!- TISSUE SPECIFICITY: PRESYNAPTIC NERVE TERMINI OF EXCITATORY
 CC SYNAPSES. IN OTHER BRAIN REGIONS, SAP97 IS FOUND IN AND ALONG
 CC BUNDLES OF UNMYELINATED AXONS. SAP97 IS NOT RESTRICTED TO THE CNS,
 CC BUT IS ALSO PRESENT AT THE BASAL LATERAL MEMBRANE BETWEEN A
 CC VARIETY OF EPITHELIAL CELLS.
 CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U14950; AAA79976.1; -.
 CC HSSP; Q12959; 1PDR.
 CC InterPro; IPR000619; Guanylate_kin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001452; SH3.
 CC Pfam; PF00018; SH3; 1.
 CC Pfam; PF00595; PDZ; 3.
 CC Pfam; PF06625; Guanylate_kin; 1.
 CC ProDom; PD000066; SH3; 1.
 CC SMART; SM00072; GUKC; 1.
 CC SMART; SM00228; PDZ; 3.
 CC SMART; SM00326; SH3; 1.
 CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 CC PROSITE; PS005052; GUANYLATE_KINASE_2; 1.
 CC PROSITE; PS0106; PDZ; 3.
 CC PROSITE; PS00002; SH3; 1.
 CC SH3 domain; Repeat. 310 PDZ 1.
 CC DOMAIN 224 404 PDZ 2.
 CC FT DOMAIN 318 404 PDZ 2.
 CC FT DOMAIN 465 545 PDZ 3.

FT DOMAIN 580 650 SH3.
 FT DOMAIN 721 911 GUANYLATE KINASE.
 FT DOMAIN 527 530 POLY-ALA.
 SQ SEQUENCE 911 AA; 100570 MW; 18CEBD31DD0CAF8B CRC64;

Alignment Scores:

Pred. No.: 6,36e-06 Length: 911
 Score: 240.50 Matches: 119
 Percent Similarity: 38.31% Conservative: 89
 Best Local Similarity: 21.92% Mismatches: 200
 Query Match: 5.84% Indels: 135
 DB: 1 Gaps: 22

US-09-502-945-4 (1-2236) x DLGL_RAT (1-911)

QY 343 AGGAAGCTGAAGAGGTGCGTGTGGACGCTGTGCACCCGAGGCGCTGGCTGAGTGTG 402
 Db 314 ArgLysAsnHisGluIleLysLeuLeuLys---GlyProLysGlyLeuGlyPheSerIle 332
 QY 403 CGTGTGGC-----CTGGAGTTGGCTGTGGCTCTTCATCTCCACCTC 447
 Db 333 AlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLysIle 352
 QY 448 ATCAAGGCGGTGAGGAGCGTGGG---CTCCAGGTAGGAGGAGCATGCTCCGG 504
 Db 353 IleGluGlyGlyAlaIleHisLysAspGlyLysLeuGlnIleGlyAspLysLeuLeuAla 372
 QY 505 ATCAATGATATTCATCTCTCTCTGATCCCATGAGGAGGTGATCAACCTCATTCGA--- 561
 Db 373 ValAsnSerValCysLeuGluGluValThrHisGluGluAlaValThrAlaLeuLysAsn 392
 QY 562 ACCAAGAAATGTGTCCATCAAGTGAGA-----CACATCGGC 600
 Db 393 ThrSerAspPheValTyrLeuLysAlaIleLysProThrSerMetTyrIleAsnAspGly 412
 QY 601 CTGATCCCGGTAAAGCTCTCCGAT---GAGCCCTCACTTGGCAGTATGTGGATCAG 657
 Db 413 TyrAlaPro-----ProAspIleThrAsnSerSerSerGlnSerValAspAsn 428
 QY 658 TTTGTGTCGAATCTGGGGGCTGCGAGGAGCGCTGGCTCCCTC----- 702
 Db 429 HisValSerProSerSerTyrLeuGlyGlnThrProAlaSerProAlaArgTyrSerPro 448
 QY 703 -----GGAAATCGGAA---AACAGGAGAGAGAGGTCTTCATCAGC 741
 Db 449 IleSerLysAlaValLeuGlyAspAspGluIleThrArgGluProArgLysValValLeu 468
 QY 742 CTGATAGCTCCGAGGCGCTTGGCTGACGATTTCCAGCGGCCCATCCAGAGCCTGGC 801
 Db 469 HisArgGlySerThrGlyLeuGlyPheAsnIleValGlyGlyGlu---AspGlyGluGly 487
 QY 802 ATCTTTATCAGCATGTG-----AAACTGGCTCCCTGTCTGTGAGGTGGGATTG 852
 Db 488 IlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu-----Leu 505
 QY 853 GAGATAGGGGACCATGTGCAAGTCAATCGGCTGAGTCTTCTTAACCTGGATCACAAG 912
 Db 506 ArgLysGlyAspArgIleIleSerValAsnSerValAspLeuAlaAlaSerHisGlu 525
 QY 913 GAGGCTGTAATGTCTGAAAAATAGCCGCGCATTCACCATCTCCATTGTAGCTGCAGCT 972
 Db 526 GlnAlaAlaAlaLeuLysAsnAla---GlyGlnAlaValThrIleValAlaGlnTyr 544
 QY 973 GGCCGGGAGCTGTTTCATGACAGACCGGGGCGGTGGCGAGGCGCGCGAG----- 1023
 Db 545 ArgProGluGluTyrSerArgPheGluAlaLysIleHisAspLeuArgGluThrMetMet 564
 QY 1024 -----CGTAGCTGCAGCGGCGAGGCTTC 1050
 Db 565 AsnSerSerValSerGlySerGlySerLeuArgThrSerGlnLysArgSerLeuTyr 584
 QY 1051 ATGCAGAGCGGCTGGCGATGGAGTCCAAACAG----- 1083


```

Db 585 ValArgAlaLeuPheAspTyrAspLysThrLysAspSerGlyLeuProSerGlnGlyLeu 604
QY 1083 ----- 1083
Db 605 AsnPhelysPheGlyAspIleLeuHisValIleAsnAlaSerAspAspGluTrpTrpGln 624
QY 1084 -----ATCCTCCAGGACGACGAG 1101
Db 625 AlaArgGlnValThrProAspGlyGluSerAspGluValGlyValIleProSerLysArg 644
QY 1102 GAGATGGAGCGGCAAGGAGAAAGAAATGTCACCAAGGACGACGAGGAATGAGAGA 1161
Db 645 ArgValGluLysLysGluArgAlaArgLeu-----LysThrValLysPheAsnSerLys 662
QY 1162 TACCGAAGGAGATGGAACAGATTAGAGGAG----- 1194
Db 663 ThrArgLysLysGlyGluIleProAspAspMetLysSerLysGlyLeuLysHisVal 682
QY 1195 -----GAAGAGAAGTTTAAAGAACAAATGGAAGAACTGGGCG 1233
Db 683 ThrSerAsnAlaSerAspSerGluSerSerTyrHisGluTyrGlyCysSerLysGlyGly 702
QY 1234 TCAAGGAACAGTACTCTTGCCTAAACCATCCTCCTCAG-----GTACAC 1281
Db 703 GlnGluGluTyrValLeuSerTyrGluProValAsnGlnGlnGluValAsnTyrThrArg 722
QY 1282 CCAGTACCCCTCGCAAGCCAAAGATATGATCAGGAGTGGAACTGAGCTCGAGCCCGCA 1341
Db 723 ProValIleIleLeuGlyProMetLysAspArgValAsnAspAspLeuIleSerGluPhe 742
QY 1342 GATGACCTGATGAGGACGAGGAGGAGGAGGAGGAGGATTCGCGGAATATGAGGAA 1401
Db 743 ProAspLysPheGlySerCysValProHisThrThrArgProLysArgAspTyrGluVal 762
QY 1402 GGCTTTGACCCCTACTCTATGTTCCACCCAGACGACATCATGGGAGGATGCCGCTC 1461
Db 763 AspGlyArgAspTyrHisPheValThrSerArgGluGlnMetGluLysAspIleGlnGlu 782
QY 1462 CTAGCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAAGGC---GGTGTGGACTCC 1518
Db 783 HisLysPheIleGluAlaGlyGlnTyrAsnAsnHisLeuTyrGlyThrSerValGlnSer 802
QY 1519 CCATTCGGAGGTGGTGGTCTTCGTGTGTATGAGCGGGGAGCTGCTGAGCGGCGATGGT 1578
Db 803 -----ValArgAlaValAlaGluLysGly-----LysHisCys 813
QY 1579 GGCATTGTGAAGGGGAGGATCATGCAATCAACGCAAGATGTGACAGACTACACC 1638
Db 814 -----IleLeuAspValSerGlyAsnAlaIleLysArgLeuGln 826
QY 1639 CTGGCTGAG 1647
Db 827 IleAlaGln 829
RESULT 13
DLG3_HUMAN STANDARD; PRT; 817 AA.
AC Q92796; O9ULI8;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic protein SAP102 (Synapse-associated protein 102).
DE (Neuroendocrine-DLG) (NE-DLG) (Discs, large homolog 3).
GN DLG3 OR KIAA1232.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=97332623; PubMed=9188857;
RA Makino K., Kuwahara H., Masuko N., Nishiyama Y., Morisaki T.,

```

```

RA Sasaki J., Nakao M., Kuwano A., Nakata M., Ushio Y., Saya H.;
RT "Cloning and characterization of NE-dlg: a novel human homolog of the
RT Drosophila discs large (dlg) tumor suppressor protein interacts with
RT the APC protein.";
RL Oncogene 14:2425-2433(1997).
RN [2]
RP SEQUENCE OF 330-817 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U49089; AAB61453.1; -.
DR EMBL; AB033058; BAA86546.1; -.
DR HSSP; Q12959; 1PDR.
DR Genew; HGNC:2902; DLG3.
DR MIM; 300189; -.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
DR SH3 domain; Repeat.
KW FT DOMAIN 130 217 PDZ 1.
FT DOMAIN 226 311 PDZ 2.
FT DOMAIN 379 465 PDZ 3.
FT DOMAIN 503 568 SH3.
FT DOMAIN 628 803 GUANYLATE KINASE.
FT CONFLICT 330 381 FTALADNHSHNSLSGLYGAVESKYSVPAPQVPVTRKSPI
FT PRHMLAEEDFT -> AARRERGAMERARKFSGSLAMGLGS
FT ASASARRASORWAMPLSLRPGGDA (IN REF. 2).
FT DFGGLSDDYVGAKNL -> SIKTKRKKSFRLSRKPFYKSK
FT ENMAQESSIQGGQVTSNTSDSESS (IN REF. 2).
SQ SEQUENCE 817 AA; 90344 MW; 3D7512EC4713FC4E CRC64;
Alignment Scores:
Pred. No.: 1.32e-05 Length: 817
Score: 234.50 Matches: 164
Percent Similarity: 37.37% Conservative: 123
Best Local Similarity: 21.35% Mismatches: 276
Query Match: 5.69% Indels: 206
DB: 1 Gaps: 35
US-09-502-945-4 (1-2236) x DLG3_HUMAN (1-817)
QY 8 CGGTGCGGTGCGGGCTCTTTCCAGCTCTGCGACGCGGGCACCC-----GAA 55

```


[illegible]

Db	418	GlyProAlaAspLeuSerGlyGlu-----LeuArgArgGlyAspArgIleLeuSerVal	433
Qy	880	AATGCGTCGCACTTCTCTAACCTGGATCACAGGAGCGCTGTAATCTGCTGAAAAATAGC	939
Db	436	AsnGlyValAsnLeuArgAsnAlaThrHisGluGlnAlaAlaAlaLeuLysArgAla	455
Qy	940	CGCAGCCTGACCATCTCCATTGTAGCTGCAGCTGGCGGAGAGCTGTTCATGCACACCGG	999
Db	456	--GlyGlnSerValThrIleValAlaGlnTyrArgProGluGluTyrSerArgPheGlu	474
Qy	1000	GAGCGGTGGCAGAGCGCGCAGCGTGCAGCTGCAG-----	1035
Db	475	SerLysIleHisAspLeuArgGluGlnMetMetAsnSerSerMetSerSerGlySerGly	494
Qy	1035	-----	1035
Db	495	SerLeuArgThrSerGluLysArgSerLeuTyrValArgAlaLeuPheAspTyrAspArg	514
Qy	1036	---CGCGAGGAGCTTCTCATGCAGAGCGCGTGGCGATCGAGCTCCAACAAGATCCTC---	1089
Db	515	ThrArgAspSerCysLeuProSerGlnGlyLeuSerPheSerTyrGlyAspIleLeuHis	534
Qy	1090	-----CAGGAGCACAGGAGATGGAGCGGCAAGG	1119
Db	535	ValIleAsnAlaSerAspAspGluTrpTrpGlnAlaArgLeuValThrProHisGlyGlu	554
Qy	1120	AGAAAGAAATT-----GCCCAGAGGCGAGAGAGAAATACAGAGATACCGG	1167
Db	555	SerGluGlnIleGlyValIleProSerLysLysArgValGluLysGluArgAlaArg	574
Qy	1168	AAGGAGATGGAACAG-----ATTGTAGAGGAGGAGAGAACTTTAAG	1209
Db	575	LeuLysThrValLysPheHisAlaArgThrGlyMetIleGluSerAsnArgAspPhePro	594
Qy	1210	AAGCAATGGGAAGAACTGGGCGCTCAAAG-----GAACAGCTACTC	1251
Db	595	GlyLeuSerAspAspTyrTyrGlyAlaLysAsnLeuLysGlyGlnGluAspAlaIleLeu	614
Qy	1252	TTGCTTAAACCATCACT---GCTGAGGTACAC-----CCAGTACCCCTTCGCAG	1299
Db	615	SerTyrGluProValThrArgGlnGluIleHisTyrAlaArgProValIleLeuGly	634
Qy	1300	CCAAAGTATGATCAGGAGTGGAACTGAGCTCGAGCCCGCAGATGACCTGGATGGAGGC	1359
Db	635	ProMetLysArgArgValAsnAspLeuIleSerGluPheProHisLysPheGlySer	654
Qy	1360	ACGAGAGACAGGAGCAGGATTTCCGGAATATGAGGAAGGCTTTCACCCCTACTCT	1419
Db	655	CysValProHisThrThrArgProArgArgAspAsnGluValAspGlyGlnAspTyrHis	674
Qy	1420	ATGTTTCCCCAGACAGATCATGGGGAAGGATGTCGGGCTCCTACGCATCAAGAAGGAG	1479
Db	675	PheValValSerArgGluGlnMetGluLysAspIleGlnAspAsnLysPheIleGluAla	694
Qy	1480	GGATCCTTAGACCTGGCCCTCGAAGGC---GGTGTGACTCCCCCATTTGGGAAGGTGTC	1536
Db	695	GlyGlnPheAsnAspAsnLeuTyrGlyThrSerIleGlnSer-----	708
Qy	1537	GTTTCTGCTGTATGACGGGAGCTGTGACGGGCAT-----GGT	1578
Db	709	ValArgAlaValAlaGluArgGly-----LysHisCysIleLeuAspValSerGly	725
Qy	1579	GGCATTTGAAAGG-----	1593
Db	726	AsnAlaIleLysArgLeuGlnGlnAlaGlnLeuTyrProIleAlaIlePheIleLysPro	745
Qy	1594	-----GACGAGATCATGGCAATCAACGGCAAGATGTGCACAGATACACCTGGCT	1644
Db	746	LysSerIleGluAlaLeuMetGluMetAsnArgArg-----GlnThrTyrGlu	761
Qy	1645	GAGCGTACGCTGCCCTGCAGAAGCGCTCG-----AATCAGGCGGAGCTGGATC	1695
Db	762	GlnAlaAsnLysIleTyrAspLysAlaMetLysLeuGluGlnGluPheGlyGluTyrPhe	781

Qy 1696 GACCTGTGGTGGTCTGCTGCCCCCAAGAGAGTATGACGATGACCTTCTTGTG 1755
Db 782 ThrAlaIleValGlnGlyAspSerLeuGluGluIleTyrAsnLysIleLysGlnIle 801
Qy 1756 AAGTCCAAAGGGGAAACAAATT 1779
Db 802 GluAspGlnSerGlyHisTyrIle 809

RESULT 14
ID DLG2_HUMAN STANDARD; PRT; 870 AA.
AC Q15700;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Channel associated protein of synapse-110 (Chapsyn-110) (Discs, large
DE homolog 2).
GN DLG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96310881; PubMed=8755482;
RA Kim E., Cho K.-O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of
RT Chapsyn-110, a member of the PSD-95 family of proteins.";
RL Neuron 17:103-113(1996).
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS AS WELL AS POTASSIUM CHANNELS.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U32376; AB04949.1; -
DR HSP; Q12959; IPDR.
DR Genew; HGNC:2901; DLG2.
DR MIM; 603583; -
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 98 184 PDZ 1.
FT DOMAIN 193 279 PDZ 2.
FT DOMAIN 421 501 PDZ 3.
FT DOMAIN 536 606 SH3.
FT DOMAIN 680 870 GUANYLATE KINASE.
SQ SEQUENCE 870 AA; 97500 MW; 89C83BA0619F6F59 CRC64;

Alignment Scores:

Pred. No.: 2.33e-05 Length: 870

Score: 230.00 Matches: 143
Percent Similarity: 35.74% Conservative: 114
Best Local Similarity: 19.89% Mismatches: 222
Query Match: 5.58% Indels: 240
DB: 32 Gaps: 1

US-09-502-945-4 (1-2236) x DLG2_HUMAN (1-870)

Qy 208 GACGTGGCCGTGCTGCTGGGAGACCTGAAGCTGGTCATCAATGAA-----CCAGCGT 261
Db 141 AspGlyArgLeuArgValAsnAspCysIleLeuArgValAsnGluValAspValSerGlu 160
Qy 262 CTGCCTCTGTTGATGCCATTCCGCCCTGCGCCCTGATCCCATCGAGCACCAGGTTGAATGAT 321
Db 161 ValSerHisSerLysAlaValGlu-----AlaLeuLysGluAlaGlySerIleAla 177
Qy 322 CAGCTGACCCCGCCGCGCTCCAGGAAGCTG-----AAGGAGGTGCGTCTGGAC 369
Db 178 ArgLeuTyrValArgArgArgProIleLeuGluThrValGluIleLysLeuPhe 197
Qy 370 CGTCTGCACCCCGGAGCGCTGCGCTGAGTGTGGTGGC-----CTG 414
Db 198 Lys---GlyProLysGlyLeuGlyPheSerIleAlaGlyValGlyAsnGlnHisIle 216
Qy 415 GAGTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGCGCGTCAGGACAGACGCTC 474
Db 217 ProGlyAspAsnSerIleTyrValThrLysIleIleAspGlyGlyAlaAlaGlnLysAsp 236
Qy 475 GGG---CTCCAGGTAGGGAGAGATCGTCGGATCAATGGATATTCATCTCCTCTGT 531
Db 237 GlyArgLeuGlnValGlyAspArgLeuMetValAsnAsnTyrSerLeuGluGluVal 256
Qy 532 ACCATGAGGAGGTCACTCACTCATTCGA---ACCAAGAAACTGTGTCCATCAAGTG 588
Db 257 ThrHisGluGluAlaValAlaIleLeuLysAsnThrSerGluValValTyrLeuLysVal 276
Qy 588 ----- 588
Db 277 GlyAsnProThrThrIleTyrMetThrAspProTyrGlyProProAspIleThrHisSer 296
Qy 589 -----AGACACATC-----GGCTGATCCCGTGAAA 615
Db 297 TyrSerProProMetGluAsnHisLeuLeuSerGlyAsnAsnGlyThrLeuGluTyrLys 316
Qy 616 AGCTCT----- 621
Db 317 ThrSerLeuProProIleSerProGlyArgTyrSerProIleProLysHisMetLeuVal 336
Qy 622 -----CCTGATGAGCCCTCACTTGGCAGATATGTGGATGATGTTGTG 663
Db 337 AspAspAspTyrThrArgProProGluProVal---TyrSerThrValAsnLysLeuLys 355
Qy 664 TCGGNATCTGGGGCGTCCGA----- 684
Db 356 AspLysProAlaSerProArgHisTyrSerProValGluCysAspLysSerPheLeuLeu 375
Qy 685 -----GGC 687
Db 376 SerAlaProTyrSerHisTyrHisLeuGlyLeuLeuProAspSerGluMetThrSerHis 395
Qy 688 AGCTGGGCTCCCTGGAAATCG----- 711
Db 396 SerGlnHisSerThrAlaThrArgGlnProSerMetThrLeuGlnArgAlaValSerLeu 415
Qy 712 GAAACACAGGACAGAGGCTTTCATCAGCTGGTAGCTCCCGAGTGTGGTGGTGGCAGC 771
Db 416 GluGlyGluProArgLysValValLeuHis---LysGlySerThrGlyLeuGlyPheAsn 434
Qy 772 ATTTCCAGCGGCCCATCCAGAACGCTGGCATCTTTATCAGCCATGTG-----AAA 822
Db 435 IleValGlyGlyGlu---AspGlyGluGlyIlePheValSerPheIleLeuAlaGlyGly 453
Qy 823 CCTGCTCCTCTGCTGAGTGGGATGGAGATGGGGACAGATTTGCGAAGTCAAT 882

Db 454 ProAlaAspLeuSerGlyGlu-----LeuGlnArgGlyAspGlnIleLeuSerValAsn 471
 QY 883 GCGTCGACTTCTTAACCTGGATCACAAGGAGGTGTAAATGCTGTAATAATAGCCGC 942
 Db 472 GlyIleAspLeuArgGlyAlaSerHisGluGlnAlaAlaAlaLeuLysGlyAla--- 490
 QY 943 AGCCTGACCATCTCATCTGACGCTGCCGGGAGCTTTCATGACACACCGGAG 1002
 Db 491 GlyGlnThrValThrIleIleAlaGlnTyrGlnProGluAspTyrAlaArgPheGluAla 510
 QY 1003 CGCTGGCAGAGCGCGCCGAC----- 1023
 Db 511 LysIleHisAspLeuArgGluGlnMetMetAsnHisSerMetSerSerGlySerGlySer 530
 QY 1024 ---CCTGAGCTCCAGCGGAGGAGCTTCTCATGCGAC----- 1056
 Db 531 LeuArgThrAsnGlnLysArgSerLeuTyrValArgAlaMetPheAspTyrAspLysSer 550
 QY 1056 ----- 1056
 Db 551 LysAspSerGlyLeuProSerGlnGlyLeuSerPheLysTyrGlyAspIleLeuHisVal 570
 QY 1057 -----AAGCGCTGGCGATGGAGTCCCAACAAG 1083
 Db 571 IleAsnAlaSerAspGluTrpTrpGlnAlaArgValMetLeuGluGlyAspSer 590
 QY 1084 -----ATCCTCCAGGAGCAGGAGATGGAGCGGCAAGAGGAAAGAA--- 1128
 Db 591 GluGluMetGlyValIleProSerLysArgArgValGluArgLysGluArgAlaArgLeu 610
 QY 1129 -----ATTGCCCAAGGAGCAGGAGGAGAAAT 1155
 Db 611 LysThrValLysPheAsnAlaLysProGlyValIleAspSerLysGlySerPheAsnAsp 630
 QY 1156 GAGAGA-----TACCGAAGGAGATGGAAACAGATGTAGAGGAGGAGAG 1200
 Db 631 LysArgLysLysSerPheIlePheSerArgLysPheProPheTyrLysAsnLysGluGln 650
 QY 1201 AAGTTTAGACAGCATGGAGAGACACTGGCGCTCAAGAACACTACTCTTGGCTAAA 1260
 Db 651 SerGluGlnGluThrSerAspProGluArgGlyGlnGluAspLeuIleLeuSerTyrGlu 670
 QY 1261 ACCATCACTGCTGAG-----GTACACCCAGTACCCCTTCGCAAGCCAAAGTAT 1308
 Db 671 ProValThrArgGlnGluIleAsnTyrThrArgProValIleIleLeuGlyProMetLys 690
 QY 1309 GATCAGGAGTGGACCTGAGCTGAGCCGCGAGATGACCTGGATGGAGCAGCAGGAG 1368
 Db 691 AspArgIleAsnAspAspLeuIleSerGluPheProAspLysPheGlySerCysValPro 710
 QY 1369 CAGGAGAGCAGGATTTCCGGAATATGAG---GAAGCTTTGACCCCTACTCTATGTC 1425
 Db 711 HisThrThrArgProLysArgAspTyrGluValAspGlyArgAspTyrHisPheValIle 730
 QY 1426 ACCCGAGACAGATCATGGGAAGGATCTCCGGCTCTTACCATCATCAAGAGGAGGATCC 1485
 Db 731 SerArgGluGln---MetGluLysAspIleGlnGluHisLysPheIleGluAlaGlyGln 749
 QY 1486 TTAGACCTGGCCCTGGAAAGG---GGTGTGACTCCCGCATTTGGGAAGGTGCTTCTT 1542
 Db 750 TyrAsnAspAsnLeuTyrGlyThrSerValGlnSer-----ValArg 763
 QY 1543 GCTGTGTATGAGCGGGAGCTGCTGAGCGCAT-----GGTGGCAT 1584
 Db 764 PheValAlaGluArgGly-----LysHisCysIleLeuAspValSerGlyAsnAla 780
 QY 1585 GTGAAGAGG----- 1593
 Db 781 IleLysArgLeuGlnValAlaGlnLeuTyrProIleAlaIlePheIleLysProArgSer 800
 QY 1594 ---GACGAGATCATGGCAATCAACGCGAGATTTGTGACAGACTACACCTGGCTGAGCT 1650
 Db 1594 ---GACGAGATCATGGCAATCAACGCGAGATTTGTGACAGACTACACCTGGCTGAGCT 1650

Db 801 LeuGluSerLeuMetGluMetAsnLysArgLeuThrGluGlnAlaLysLysThrTyr 820
 QY 1651 GACGCTGCCGCGAGAGCGCTGGAATCGACGGCGGACATCGACCTGTGTGTT 1707
 Db 821 AspArgAlaIle---LysLeuGluGlnGluPheGlyGluTyrPheThrAlaIleVal 838
 RESULT 15
 DLG2_RAT
 ID DLG2_RAT STANDARD; PRT; 852 AA.
 AC Q63622; Q62939; P70548;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Channel associated protein of synapse-110 (Chapsyn-110) (Synaptic
 DE density protein PSD-93) (Discs, large homolog 2).
 GN DLG2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96310881; PubMed=8755482;
 RA Kim E., Cho K.O., Rothschild A., Sheng M.;
 RT "Heteromultimerization and NMDA receptor-clustering activity of
 RL Chapsyn-110, a member of the PSD-95 family of proteins.";
 RN Neuron 17:103-113(1996).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96193770; PubMed=8625413;
 RA Brennan J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E.,
 RA Santillan D.R., Wu Z., Huang F., Xia H., Peters M.F.,
 RA Froehner S.C., Bredt D.S.;
 RT "Interaction of nitric oxide synthase with the postsynaptic density
 RL protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
 RN Cell 84:757-767(1996).
 [3]
 RP SEQUENCE FROM N.A.
 RX Irie M., Hata Y., Takai Y.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS AS WELL AS POTASSIUM CHANNELS.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U49049; AAB53243.1; -;
 CC EMBL; U50717; AAC52643.1; -;
 CC EMBL; U53368; AAB48562.1; -;
 CC HSPS; O12959; IPRD;
 CC InterPro; IPR000619; Guanylate_kin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001452; SH3.
 CC Pfam; PF00018; SH3; 1.
 CC Pfam; PF00595; PDZ; 3.
 CC Pfam; PF00625; Guanylate_kin; 1.
 CC ProDom; PD000066; SH3; 1.
 CC SMART; SM00072; GUKC; 1.
 CC SMART; SM00228; PDZ; 3.
 CC SMART; SM00326; SH3; 1.
 CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 CC PROSITE; PS00852; GUANYLATE_KINASE_2; 1.
 CC PROSITE; PSS0106; PDZ; 3.
 CC PROSITE; PSS0002; SH3; 1.

SH3 domain; Repeat. 184 PDZ 1.
 FT DOMAIN 98 184 PDZ 2.
 FT DOMAIN 193 279 PDZ 3.
 FT DOMAIN 421 501 SH3.
 FT DOMAIN 536 606
 FT DOMAIN 662 852 GUANYLATE KINASE.
 FT CONFLICT 181 182 VR -> IL (IN REF. 2).
 FT CONFLICT 228 228 I -> M (IN REF. 2).
 FT CONFLICT 326 326 R -> K (IN REF. 2).
 FT CONFLICT 339 339 D -> E (IN REF. 3).
 FT CONFLICT 450 454 MISSING (IN REF. 2).
 FT CONFLICT 464 465 GD -> RK (IN REF. 2).
 FT CONFLICT 474 474 D -> H (IN REF. 2).
 FT CONFLICT 476 476 R -> P (IN REF. 2).
 FT CONFLICT 478 478 A -> D (IN REF. 2).
 FT CONFLICT 484 486 AAA -> LP (IN REF. 2).
 FT CONFLICT 506 506 A -> S (IN REF. 2).
 FT CONFLICT 569 569 H -> N (IN REF. 2).
 FT CONFLICT 586 586 L -> Q (IN REF. 2).
 FT CONFLICT 626 641 MISSING (IN REF. 2).
 FT CONFLICT 639 639 K -> A (IN REF. 3).
 FT CONFLICT 726 726 F -> L (IN REF. 1).
 FT CONFLICT 733 733 N -> Y (IN REF. 2).
 FT CONFLICT 749 749 E -> V (IN REF. 1).
 FT CONFLICT 756 756 L -> H (IN REF. 2).
 FT CONFLICT 791 792 KR -> NG (IN REF. 2).
 FT CONFLICT 794 794 T -> M (IN REF. 2).
 SQ SEQUENCE 852 AA; 94934 MW; F8D414A8B9CF5B09 CRC64;

Alignment Scores:

Pred. No.: 2.63e-05 Length: 852
 Score: 229.00 Matches: 145
 Percent Similarity: 36.09% Conservative: 108
 Best Local Similarity: 20.68% Mismatches: 226
 Query Match: 5.56% Indels: 222
 DB: 1 Gaps: 33

US-09-502-945-4 (1-2236) x DLG2_RAT (1-852)

QY 208 GAGTGGCGCTCTCTGGGAGACCTGAAGCTGGTCAATCAATGAA-----CCGAGCGCT 261
 DB 141 AspGlyArgLeuArgValAsnAspCysIleLeuArgValAsnGluValAspValSerGlu 160
 QY 262 CTGCTCTCTTTGATCCATCTGGCGCTGATCCCACTGAAGCACCAGCTGGAATATGAT 321
 DB 161 ValSerHisSerLysAlaValGlu-----AlaLeuLysGluAlaGlySerIleVal 177
 QY 322 CAGCTGACCCCGCGCTCCAGGAAGCTG-----AAGAGTGGCTCTGGAC 369
 DB 178 ArgLeuTyrrValArgArgArgProIleLeuGluThrValValGluIleLysLeuPhe 197
 QY 370 CGTCTCACCCGAGGCTCGGCTCGAGTGGCTGGTGGC-----CTG 414
 DB 198 Lys---GlyProLysGlyLeuGlyPheSerIleAlaGlyValGlyAsnGlnHisIle 216
 QY 415 GAGTTTGGTGGCTCTTCATCTCCACCTCATCAAGCGGTCAGGCGGTCAGCAGACGCTC 474
 DB 217 ProGlyAspAsnSerIleTyrrValThrLysIleIleAspGlyGlyAlaAlaGlnLysAsp 236
 QY 475 GGG---CTCCAGTAGGGGAGAGATCGCCGGATCAATGATATTCATCTCTCTCTCT 531
 DB 237 GlyArgLeuGluValGlyArgLeuLeuMetValAsnAsnTyrrSerLeuGluGluVal 256
 QY 532 ACCCATGAGGAGTCTACACCTCATTCGA---ACCAAGAAACTGTGTCATCAAGTG 588
 DB 257 ThrHisGluGluAlaAlaIleLeuLysAsnThrSerAspValValTyrrLeuLysVal 276
 QY 588 ----- 588
 DB 277 GlyLysProThrThrIleTyrrMetThrAspProTyrrGlyProProAspIleThrHisSer 296
 QY 589 -----AGACACATC-----GGCTGATCCCGGTGAA 615
 DB 589 ----- 615

Db 297 TyrSerProProMetGluAsnHisLeuLeuSerGlyAsnAsnGlyThrLeuGluTyrrLys 316
 QY 616 AGCTCT----- 621
 Db 317 ThrSerLeuProProIleSerProGlyArgTyrrSerProIleProLysHisMetLeuVal 336
 QY 622 -----CCTGATGAGCCCTCACTTGGCAGATGTGGATCATGTTTGTG 663
 Db 337 GluAspAspTyrrThrArgProGluProVal---TyrSerThrValAsnLysLeuCys 355
 QY 664 TCGGAATCTGGGGCGTCCGA----- 684
 Db 356 AspLysProAlaSerProArgHisTyrrSerProValGluCysAspLysSerPheLeuLeu 375
 QY 685 -----GGC 687
 Db 376 SerThrProTyrrProHisTyrrHisLeuGlyLeuLeuProAspSerAspMetThrSerHis 395
 QY 688 AGCTGGCTCCCTCGAATCGG----- 711
 Db 396 SerGlnHisSerThrAlaThrArgGlnProSerValThrLeuGlnArgAlaIleSerLeu 415
 QY 712 GAAACAGGAGAGAGAGTCTTCATCAGCCTGCTAGGCTCCGAGGCTTGGCTGCAGC 771
 Db 416 GluGlyGluProArgLysValLeuHis---LysGlySerThrGlyLeuGlyPheAsn 434
 QY 772 ATTTCCAGCGCCCTCCAGAGCCTGGCATCTTTATCAGCCATGTG-----AAA 822
 Db 435 IleValGlyGlyGlu---AspGlyGlyIlePheValSerPheIleLeuAlaGlyGly 453
 QY 823 CTGCTCCCTCTCTGCTGAGTGGGATGGATGGGACCATGTCGATGTCAGTCAAT 882
 Db 454 ProAlaAspLeuSerGlyGlu-----LeuGlnArgGlyAspGlnIleLeuSerValasn 471
 QY 883 GCGCTCGACTTCTTAACCTGGATCACAAGGAGCTGTAATGTCTGAAAATAGCCGC 942
 Db 472 GlyIleAspLeuArgGlyAlaSerHisGlnAlaAlaAlaLeuLysGlyAla--- 490
 QY 943 AGCTGACCATCTCCATTTAGTGCAGCTGGCGGAGAGCTGTTTCATGACAGCGGAG 1002
 Db 491 GlyGlnThrValThrIleIleAlaGlnTyrrGlnProGluAspTyrrAlaArgPheGluAla 510
 QY 1003 CGGCTGGCAGAGCGCGGCAG----- 1023
 Db 511 LysIleHisAspLeuArgGluGlnMetMetAsnHisSerMetSerSerGlySerGlySer 530
 QY 1024 ---CGTAGCTGCAGCGCAGAGCTTCTCATGCAGAGAGCGGCTGGCAGTGGATGCCAAC 1080
 Db 531 LeuArgThrAsnGlnLysArgSerLeuTyrrValArgAlaMetPheAspTyrrAspLysSer 550
 QY 1081 AAG-----ATC 1086
 Db 551 LysAspSerGlyLeuProSerGlnGlyLeuSerPheLysTyrrGlyAspIleLeuHisVal 570
 QY 1087 CTCAGGAGCAGCAGAGATGGAGCGGCAAGAGAGAAAGAAAT----- 1131
 Db 571 IleAsnAlaSerAspGluTyrrPrgLlnAlaArgValIleLeuAspGlyAspSer 590
 QY 1132 -----GCCAGAGGAGCAGCAGAGAGAAATGAGATACCGAGAG 1170
 Db 591 GluGluMetGlyValIleProSerLysArgValGluArgLysGluArgAlaArgLeu 610
 QY 1171 GAGATGGAGACAG-----ATTGTAGAGGAGAGAGAGAGTTTAAAGAG 1212
 Db 611 LysThrValLysPheAsnAlaLysProGlyValIleAspSerLysGlyAspIleProGly 630
 QY 1213 CAATGGGAGAGAGATGGGCTCAAG-----GAACAGCTACTCTTGCT 1257
 Db 631 LeuGlyAspGlyTyrrGlyThrLysThrLeuArgGlyGlnGluAspLeuLeuSer 650
 QY 1258 ---AAACCATCATCTCTGAG-----GTACACCCAGTACCCCTCGCAAGCA 1302
 Db 651 TyrGluProValThrArgGlnGlnIleAsnTyrrThrArgProValIleIleLeuGlyPro 670

Search completed: March 21, 2003, 12:48:08
Job time : 53.2735 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:20:42 ; Search time 76.6811 Seconds
(without alignments)
12016.549 Million cell updates/sec

Title: US-09-502-945-4
Perfect score: 4120
Sequence: 1 cctgcccgcgcgcgtgcg.....tattttccagcttaaaaaa 2236

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p.model -DEV=xlx
-Q=/cgn2_1/USPRO_pool/US09502945/runat_14032003_101058_19113/app_query.fasta_1.10979
-DB=SPTREMBL_21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1565 -runat_14032003_101058_19113 -NCPU=6 -ICPU=3
-NO_XLPAY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2720.5	66.0	533	4 Q96B29	Q96B29 homo sapien

ID	Q96B29	PRELIMINARY;	PRT;	533 AA.
AC	Q96B29;			
DT	01-DEC-2001 (Tremblrel. 19, Created)			
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)			
DE	Similar to PDZ-73 protein.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLON;			
RA	Strausberg R.;			
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC016057; AAH16057.1;			
DR	InterPro: IPR001478; PDZ.			
DR	PFam: PF00595; PDZ; 3.			
DR	PROSITE: PS0106; PDZ; 3.			
SQ	SEQUENCE 533 AA; 60314 MW; A9D64A3B011E2FBD CRC64;			

Alignment Scores:
Pred. No.: 1.4e-179 Length: 533

Score: 2720.50 Matches: 532
Percent Similarity: 96.55% Conservative: 0
Best Local Similarity: 96.55% Mismatches: 0
Query Match: 66.03% Indels: 19
DB: 4 Gaps: 1

US-09-502-945-4 (1-2236) x Q96B29 (1-533)

QY	97	ATGGACCGAAAGTGGCCCGGAGAAATTCGGGATAGGTGGATTTCTGATGAAATGAT	156
DB	1	MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp	20
QY	157	GCAGAGAGGACTATCTCTATGATGTGTCGGAATGTACCACGACCATGACGTGCC	216
DB	21	AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla	40
QY	217	GTGCTGCTGGGAGACTGAAGCTGTCATCAATGAACCCAGCCGCTGCTCTGTTGAT	276
DB	41	ValIleValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp	60
QY	277	GCCATTGGCCCGCTGATCCACTGAAGCACCAGGTGGAATATGATCAGCTGACCCCGGG	336
DB	61	AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg	80
QY	337	CGCTCCAGGAGCTGAAGAGGTGCTGTGGACCGTCTGCACCCCGAGAGGCTCGGCCTG	396
DB	81	ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu	100
QY	397	ACTGTGCTGCTGGCTGGCTGAGTGTGGCTGTGGCTTTCATCTCCACCTCATCAAGGC	456
DB	101	SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly	120
QY	457	GTTCAGGCAGACAGCTGGGCTGCAGGTAGGGAGCAGATCTCGGATCAATGGAT	516
DB	121	GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr	140
QY	517	TCCATCTCTCTCTGATCCCATGAGGAGTTCATCAACTCATTCGAACCAAGAAACTGTG	576
DB	141	SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysThrVal	160
QY	577	TCCATCAAGTGAACATCGGCTGATCCCGTGAAGTCTCTCATGAGCCCTC	636
DB	161	SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu	180
QY	637	ACTTGGCAGTATGTCAGTGTGTGTCGGAATCTGGGAGTCTGGGAGCAGCTGGGC	696
DB	181	ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly	200
QY	697	TCCCTCGAAATCGGAAACAGAGAGAAAGGTCTTCATCAGCTGTGAGCTCCCGA	756
DB	201	SerProGlyAsnArgGluAsnLysGluLysLysValPheIleSerLeuValGlySerArg	220
QY	757	GGCTTGGCTGACGATTTCCACGGCCCATCCAGAGCTGGCATCTTATCAGCCAT	816
DB	221	GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis	240
QY	817	GTCAACCTGGCTGCTGCTGCTGAGTGGGATGTGAGATAGGGACGAGATGTGCGAA	876
DB	241	ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu	260
QY	877	GTCAATGGCTGCTGCTGCTGCTGAGTCAACAGAGGCTGTAATGTGCTGAAAAAT	936
DB	261	ValAsnGlyValAspPheSerAsnLeuAspHisLysGlu	273
QY	937	AGCGGAGCTGACCATCTCTCCATGTAGTCGAGCTGGCGGAGCTGTTCATGACAGAC	996
DB	274	-----GlyArgGluLeuPheMetThrAsp	281
QY	997	CGGAGCGGCTGGCAGAGCGCGGACCGTGTGAGCTGCGCGGAGGAGCTTCATGCGAG	1056
DB	282	ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln	301
QY	1057	AAGCGGCTGCGAGTGGAGTCCAAACAGATCTCCAGGAGCAGGAGATGGAGCGGCA	1116

DB	302	LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln	321
QY	1117	AGGAGAAAGAAATTTGCCAGAGGAGGAGAGAGAAATGAGAGATACCCGAGAGAGATG	1176
DB	322	ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet	341
QY	1177	GAACAGATTGTAGAGGAGGAGAGAGTTTAAAGCAATGGGAAGAGACTGGGGCTCA	1236
DB	342	GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer	361
QY	1237	AAGGAACAGCTACTTCTTCCCTAAAACCATCCTCCTCAGGTACACCCAGTACCCCTTCGC	1296
DB	362	LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg	381
QY	1297	AAGCAAAAGTATGATCAGGAGTGGAACTGAGCTCGAGCCCGCAGATGACCTGGATGA	1356
DB	382	LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspAspLeuAspGly	401
QY	1357	GGCAGGAGGAGGAGGAGGAGGAGGATTCGGGAAATATGAGGAAGCTTTGACCCCTAC	1416
DB	402	GlyThrGluGluGlnGlyGluGlnAspPheArgLysTyrGluGluGlyPheAspProTyr	421
QY	1417	TCTATGTTCACTCCAGAGCAGATCATGGGAAGGATGTCCGGCTCTCCTACGATCAAGAAG	1476
DB	422	SerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLys	441
QY	1477	GAGGATCTCTTAGACCTGGCCCTGGAAGGGCTGTGACTCCCCCATTTGGGAAGGTGTC	1536
DB	442	GluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProIleGlyLysValVal	461
QY	1537	CTTTCTGCTGTATGAGCGGGGAGCTGCTGAGCGGATGTGCGCATTTGTAAGGGGAC	1596
DB	462	ValSerAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAsp	481
QY	1597	GAGATCATGGCAATCAAGCAAGATTGTGACAGCTACACCTGGCTGGCTGAGGCTGACGCT	1656
DB	482	GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAla	501
QY	1657	GCCCTGCAAGAGGCTGGAATCAGGGGGGAGCTGATCGACCTGGATCGACCTGGCTGCGCTGC	1716
DB	502	AlaLeuGlnLysAlaAlaTrpAsnGlnGlyLysAspTrpIleAspLeuValValAlaValCys	521
QY	1717	CCCCAAAGGAGTATGACGATGACCTGACCTTC 1749	
DB	522	ProProLysGluTyrAspAspGluLeuThrPhe 532	

RESULT 2

Q9ES65	PRELIMINARY;	PRT;	548 AA.
AC	Q9ES65;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	Harmonin isoform al.		
GN	2010016F01RIK OR USH1C.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20428180; PubMed=10973247;		
RA	Verpy E., Leibovici M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,		
RA	Mansour A., Blanchard S., Kobayashi I., Keats B.J., Slim R., Petit C.;		
RT	"A defect in harmonin, a PDZ domain-containing protein expressed in		
RT	the inner ear sensory hair cells, underlies Usher syndrome type 1C.";		
RL	Nat. Genet. 26:51-55(2000).		
DR	EMBL; AF228924; AAG12457.1;		
DR	HSSP; P29476; 1QAV.		
DR	MGI; 1919338; 2010016F01RIK.		
DR	InterPro; IPR001478; PDZ.		
DR	Pfam; PF00595; PDZ; 3.		

the inner ear sensory hair cells, underlies Usher syndrome type 1C.;

RL Nat. Genet. 26:51-55(2000).

DR EMBL; AF228925; AAG12458.1; ..

DR HSP; P29476; IQAV

DR MGD; MGI:1919338; 2010016P01R1k.

DR InterPro; IPR001478; PDZ.

DR InterPro; IPR002965; P_Rich_extensn.

DR Pfam; PF00595; PDZ; 3.

DR PRINTS; PR01217; PRICEXTENS.

DR SMART; SM00228; PDZ; 3.

DR PROSITE; PS0106; PDZ; 3.

SQ SEQUENCE 910 AA; 102284 MW; 427B97953BA5D941 CRC64;

Alignment Scores:

Pred. No.: 2,04e-161 Length: 910

Score: 2459.00 Matches: 535

Percent Similarity: 59.62% Conservatave: 23

Best Local Similarity: 57.16% Mismatches: 50

Query Match: 59.68% Indels: 328

DB: 11 Gaps: 10

US-09-502-945-4 (1-2236) x Q9ES64 (1-910)

QY 97 ATGGACCGAAAGTGGCCGCGAGAAATCCGGCATAAGGTGGATTTCTGATTGAAATGAT 156

DB 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20

QY 157 GCAGAGAGGACTACTCTATGATGTGCTGCGAATGTACCACAGACCATGAGCTGGCC 216

DB 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40

QY 217 GTGCTCGTGGGAGACCTGAAGTGTGTCATCAATGAACCCAGCCGTCGCTCTGTTTCAT 276

DB 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProAsnArgLeuProLeuPheAsp 60

QY 277 GCCATTCCGGCGCTGATCCCACTGAAGCACCAGGTGGAAATGATGACCTGACCCCGCCGG 336

DB 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80

QY 337 CGCTCCAGAGAGCTGAAGAGGTGGCTGCGACCGTTCGACCCGCGAGGGCTCGGCCGTG 396

DB 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100

QY 397 AGTGTGCTGTGGCTGGAGTTGGCTGTGGCTGCTTCATCTCCACCTCATCAAGGC 456

DB 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120

QY 457 GGTACGGCAGACAGCTCGGCTCCAGTAGGGAGAGATCGTCCGGATCAATGGATAT 516

DB 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140

QY 517 TCCATCTCCTCTGTACCATGAGGAGGTTCATCACTCATTCGACCAACCAAGAACTGTG 576

DB 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160

QY 577 TCCATCAAGTGAGACATCGGCTGTATCCCGTGAAGCTCTCTGATGAGCCCTC 636

DB 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProGluGluSerLeu 180

QY 637 ACTTGGCAGTAGTGGATTCAGTTTGTGCGGAATCTGGGGCGGTGGAGGAGCGCTGGCC 696

DB 181 LysTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlyGlyLeuGly 200

QY 697 TCCCTCGGAATCGGGAACAAGAGAGAGTCTTTCATCAGCCTGTGTAGGCTCCCGA 756

DB 201 SerProGlyAsnArgThrThrLysGluLysLysValPheIleSerLeuValGlySerArg 220

QY 757 GGCCTTGGCTGCAGCATTTCCAGGGGCCCATCCAGAGCCTGGCATCTTTATCAGCCAT 816

DB 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheValSerHis 240

QY 817 GTGAACCTGGCTCCCTGCTGCTGAGTGGGATGGAGATAGGGACCATGTCGAA 876

DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluThrGlyAspGlnIleValGlu 260

QY 877 GTCAATGGCGTCGACTTCTCTAAGTGCATCACAAGAGGCTGTPAAATGTGCTGAAAT 936

DB 261 ValAsnGlyIleAspPheThrAsnLeuAspHisLysGluAlaValAsnValLeuLysSer 280

QY 937 AGCCGACGCTGACCATCTCATGTAGCTGACCTGCGGGAGCTGTTTCATCAGACAG 996

DB 281 SerArgSerLeuThrIleSerIleValAlaGlyArgGluLeuPheMetThrAsp 300

QY 997 CGGAGGCGGTGGCAGAGCGCGCAGCGTGCAGCTGCAGCGGAGGAGCTCTCATCGAG 1056

DB 301 ArgGluArgLeuGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320

QY 1057 AAGCGGTGGCGATGGAGTCCACAAGATCTCCAGAGCAGCAGAGATGGAGCGCAA 1116

DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnMetGluArgGln 340

QY 1117 AGGAGAAAAGAAATGCCAGAGGCGCAGAGGAAATGAGAGATACCGGAGGAGATG 1176

DB 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360

QY 1177 GAACAGATTGTAGAGGAGGAGAGAGTAAAGCAATGGGAAGACTGGGGCTCA 1236

DB 361 GluGlnIleSerGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380

QY 1237 AAGAACAGCTACTCTTCCTTAAACCATCAGTCTGAGTACACCCAGTACCCCTTCGC 1296

DB 381 LysGluGlnLeuIleLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400

QY 1297 AAGCCA----- 1302

DB 401 LysProLysSerPheGlyTyrPheTyrArgTyrAspGlyLysPheProThrIleArgLys 420

QY 1303 -----AAGTATGAT----- 1311

DB 421 LysAlaLysGluLysLysLysAlaLysTyrAspSerLeuGlnAspLeuArgLysAsnLys 440

QY 1311 ----- 1311

DB 441 LysGluLeuGluPheGluGlnLysLeuTyrLysGluLysGluMetLeuGluLysGlu 460

QY 1311 ----- 1311

DB 461 LysGlnLeuLysIleAsnArgLeuAlaGlnGluValSerGluThrGluArgGluAspLeu 480

QY 1311 ----- 1311

DB 481 GluGluSerGluLysThrGlnTyrTrpValGluArgLeuCysGlnThrArgLeuGluGln 500

QY 1312 -----CAGGAGTGGAGACCTGACCTC----- 1332

DB 501 IleSerSerAlaGluAsnGluIleProGluMetThrThrGlyProProProProPro 520

QY 1332 ----- 1332

DB 521 SerValSerProLeuAlaProProLeuArgArgPheAlaGlyIleHisLeuHisThr 540

QY 1332 ----- 1332

DB 541 ThrAspLeuAspAspIleProLeuAspMetPheTyrTyrProLysThrProSerAla 560

QY 1332 ----- 1332

DB 561 LeuProValMetProHisProProSerValAsnSerProSerLysValProAlaProPro 580

QY 1332 ----- 1332

DB 581 ValLeuProSerSerGlyHisValSerSerSerSerProTrpValGlnArgThrPro 600

QY 1332 ----- 1332

DB 601 ProProIleProLeuProProProProSerIleProThrGlnAspLeuThrProThrArg 620

RESULT 6
 Q9W443
 ID Q9W443 PRELIMINARY: PRT; 493 AA.
 AC Q9W443;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CG5921 protein.
 GN CG5921.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-F., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "the genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003436; AAF46116.1; -;
 DR HSSP; Q12959; 1PDR.
 DR FlyBase; FBgn0029835; CG5921.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 2.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 SQ SSEQUENCE 493 AA; 50574 MW; C46711C9F19797A4 CRA64;
 Alignment Scores:
 Pred. No.: 5,31e-21 Length: 493
 Score: 425.50 Matches: 128
 Percent Similarity: 42.76% Conservative: 58
 Best Local Similarity: 29.43% Mismatches: 164
 Query Match: 10,33% Indels: 85
 DB: 5 Gaps: 11


```
QY 595 ATCCGCTGATCCCGTGAAGAGCTCTCTCTGATGAGCCCTCCTACCTGTCGATGATGATGATGATGATGAT 654
Db 221 AlaGlyArgIleProGlyGlyPyrValThrAsnHisIleTyrThrTriP-----ValAsp 238
QY 655 -----CAGTTTGTGTCGGAATCTGGGCGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 696
Db 239 ProGlnGlyArgSerIleProSerGlyLeuProGlnProHisGlyGlyAlaLeu 258
QY 697 -----TCCCTCGGAATCGGGAAC-----AAGGAGAG 726
Db 259 ArgGlnGlnGlyAspArgSerThrLeuHisLeuGlnGlyGlyAspGluLys 278
QY 727 AAGGCTTCATCAGCTGATGAGCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 786
Db 279 LysValAsnLeuValLeuGlyArgSerGlyArgSerLeuGlyLeuThrIleArgGlyGlyAla 298
QY 787 ATCCAGAACCTCGGCTCTTATCAGCCATGTGAACCTGGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 846
Db 299 GluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySer 318
QY 847 GGATTGGAGATAGGGACAGAGATTGTGCAAGTCAATGGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 906
Db 319 GlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeu 338
QY 907 CACAGGAGGCTGTAATCTGCTGAAATAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 966
Db 339 HisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr-----ValLys 357
QY 967 GCAGCTGGCGCG-----GAGCTGTTTCATCAGACAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1014
Db 358 AspValGlyArgLeuProHisAlaArgThrThrValAspGluThrIleAlaSer 377
QY 1015 GCGCGCAGCGTGAAGTGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1074
Db 378 SerArgIleArgGluThrMetAlaAsnSerAlaGlyPheLeuGlyAspLeuThrThrGlu 397
QY 1074 ----- 1074
Db 398 GlyIleAsnLysProGlyPheTyrLysGlyProAlaGlySerGlnValThrLeuSerSer 417
QY 1075 -----TCCAAAGATCTCCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1125
Db 418 LeuGlyAsnGlnThrArgValLeuLeuGluGluGluAlaArgHisLeuLeuAsnGluGln 437
QY 1126 GAAATGGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1167
Db 438 GluHisAlaThrMetAlaTyrTyrLeuAspGlu--TyrArgGlyGlySerValSerVal 456
QY 1168 -----AAGGAGATGGAACAGATGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1215
Db 457 GluAlaLeuValMetAlaLeuPheLysLeuLeuAsnThrHisAlaLysPheSerLeuLeu 476
QY 1216 TGGGAA-----GAGACTGGGCTCAAGGAGCAGCTACTCTTG 1254
Db 477 SerGluValArgGlyThrIleSerProGlnAspLeuGluArgPheAspHisLeuValLeu 496
QY 1255 CCTAAACCATCAGTCTGAGGTACCCAGTACCCCTTCGCAAGCCA----- 1302
Db 497 ArgArg-----GluIleGluSerMetLysAlaArgGlnProGlyProGly 512
QY 1303 -----AAGTAT-----GATCAG 1314
Db 513 AlaGlyAspThrTyrSerMetValSerTyrSerAspThrGlySerSerThrGlySerHis 532
QY 1315 GGAGTGAACCTCAGCTCGAGCCCGCA--GATGACTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1371
Db 533 GlyThrSerThrThrValSerSerAlaArgAsnThrLeuAsp-----LeuGluGluThr 550
QY 1372 GGAGAG-----CAGGATTTCGG 1389
Db 551 GlyGluAlaValGlnGlyAsnIleAsnAlaLeuProAspValSerValAspValArg 570
QY 1390 AAATATGAGGAGGCTTTGACCCCTACTCTATGTTCTC-----ACCCGAGCAGCAGATCATG 1443
```

```
Db 571 SerThrSerGlnGlyLeuSerSerPheLysProLeuProArgProProLeuAlaGln 590
QY 1444 GGAAGAGATGTCGGCTCTACGCATCAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1503
Db 591 GlyAsnAspLeuProLeuGlyGlnProArgLysLeuGlyArgGluAspLeuGlnProPro 610
QY 1504 GCGGCTGTGGACTCCCACTGGGAAGGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1563
Db 611 SerSerThrProSerCysSerGly--ThrValPheSerAlaProGlnAsnArgSerPro 629
QY 1564 CTGAGCGGCATGTGGCATTTGTAAAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1623
Db 630 Pro-AlaGly----- 632
QY 1624 GTGACAGACTACACCTCGCTGAGCTGACGCTGCCCTGCAGAGAGGCTGGAATCAGGCG 1683
Db 633 -----ThrAlaPro-----ThrPr 637
QY 1684 GGGAGCTGGATCGACCTTGTGTTGCGCTGTGCCCCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1743
Db 637 oGlyThrSerSerAlaGlnAspLeuProSerSerLys----- 649
QY 1744 ACCTCTCTGCTGAAGTCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1803
Db 650 --IleTyrAlaSerValSerProAlaAsnProSerSerLys----- 662
QY 1804 CGGCCCCACCTCGTGAACACAAAGCCTCGGACGAGCCTTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1863
Db 663 -----ArgProLeuAspAlaH 668
QY 1864 ACCAGATGGCATCTTGGGAGCTGAATCTATCATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1923
Db 668 Is-----LeuAlaLeuValAsnGlnHisProIleGlyP 679
QY 1924 CTGAACCA 1931
Db 679 roPhePro 681
RESULT 9
Q9H5P4 PRELIMINARY; PRT; 517 AA.
AC Q9H5P4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CDNA: FLJ23209 fis, clone ADH00512 (Hypothetical 55.7 kDa protein).
DE protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADIPOSE TISSUE;
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026862; BAB15577.1; -
DR EMBL; BC015692; AALH5692.1; -
DR HSSP; Q12923; 3PDZ.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
KW Hypothetical protein.
```



```

SQ SEQUENCE 517 AA: 55676 MW: 06C9C614283478F7 CRC64;

Alignment Scores:
Pred. No.: 3,52e-11 Length: 517
Score: 283.50 Matches: 104
Percent Similarity: 42.71% Conservative: 63
Best Local Similarity: 26.60% Mismatches: 162
Query Match: 6.88% Indels: 62
DB: 4 Gaps: 11

US-09-502-945-4 (1-2236) x Q9H5P4 (1-517)
Qy 349 CTGAAGAGGTCGTCGTGACCGCTGCACCCCGGAGGCTCGGCTGAGTGTGCTGGT 408
Db 84 lIeHISerValArgValGluLysSerProAlaGlyArgLeuGlyPheSerValArgGly 103
Qy 409 GGCCTGGAGTTGGCTGGGCTCTTCATCTCCACCTCATCAAAAGGGGTCAGCAGAC 468
Db 104 GlySerGluHisGlyLeuGlyIlePheValSerLysValGluGluGlySerAlaGlu 123
Qy 469 AGCGTCGGGCTCCAGTAGGAGGAGGATCGTCGGATCAATGGATATTCATCTCTCC 528
Db 124 ArgAlaGlyLeuGlyValGlyLysPheValSerLysValGluGluGlySerLeuGluSer 143
Qy 529 TGTACCCATGAGGAGGCTCATCAACCTCATTCGAACCAAGAAACCTGTCTCCATCAAGTG 588
Db 144 ThrThrMetGlySerAlaValLysValLeuThrSerSerArgLeuHisMetMetVal 163
Qy 589 AGACATCGGCTGATCCCGGTGAAGAGCTCTCTGATGAGCCCTCATCTGGCAGTAT 648
Db 164 ArgArgMetGlyArgValProGlyIleLysPheSerLysGluLysThrTrp----- 181
Qy 649 GTGGAT-----CAGTTTCTGTCGGATCTGGGCGCTCGGAGGAGCGCTGGGC 696
Db 182 ValAspValAlaSerArgLeuValGluLys-----CysGlySerThrPro 198
Qy 697 TCCCTCGGAATCGGGAACCAAGAGAGAGGCTTCATCAGCGCTGGTGGCTCCCGA 756
Db 199 SerAspThrSerSerGluAspGlyValArgAlaGlyIleValHisLeuThrThrSerAsp 218
Qy 757 GGC-----CTTGGCTGAGCATTTCCAGCGGCCCATCCAGAAAGCGCTGTCATTC 810
Db 219 AspPheCysLeuGlyPheAsnIleArgGlyLysGluPheGlyLeuGlyIleThrVal 238
Qy 811 AGCCATGTGAACCTGGCTCCTCTGCTGAGTGGGATTCGAGATAGGGAGCAGGAT 870
Db 239 SerLysValAspHisGlyGlyLeuAlaGluGluAsnGlyIleLysValGlyAspGlnVal 258
Qy 871 GTCGAAGTCAATGGCTCGACTCTCTTAACCTGGATCAAGAGGCTGTAATGTGCTG 930
Db 259 LeuAlaAlaAsnGlyValArgPheAspIleSerHisSerGlnAlaValGluValLeu 278
Qy 931 AAAATAGCCGAGCTGACCATCTCCATTTAGCTGAGTGGGATTCGAGATAGGGAGCAGT 990
Db 279 LysGlyGlnThrHisIleMetLeuThrIle---LysGluThrGlyArg-----TyrPro 295
Qy 991 ACAGACGGGAGCGCTCGCAGAGCGCGGAGCGTGCAGCTGAGCTGCACCGCAGAGCTTCTC 1050
Db 296 AlaTyrLysGluMetValSerGluTyrCysTrpLeuAspArgLeuSerAsnGlyValLeu 315
Qy 1051 ATGCAGACGGCTGGCGATGGATGCCAACAGATCTCCAGGAGCAGCAGGAGATGGAG 1110
Db 316 GlnGlnLeuSerProAlaSerGluSerSerSerSerValSerSerCysAlaSerSerAl 335
Qy 1111 CGGCAAGAGAGAAAGAAATTCGCCAGAGGCGAG-----CAGAGGAA 1152
Db 335 aProTyrSerSerGlySerLeuProSerAspArgMetAspIleCysLeuGlyGlnGlu 355
Qy 1153 AATGAGATACCGGAAGAGATGGAACAGA----- 1183
Db 355 uProGlySerArgGlyProGlyTrpGlyArgAlaAspThrAlaMetGlnThrGluProAs 375
Qy 1184 -----TTGTAGAGGAGGAGAG 1200

Db 375 pAlaGlyGlyArgValGluThrTrpCysSerValArgProThrValIleLeuArgAspTh 395
Qy 1201 AAGTTTAAGAACCAATGGGAAGAAGACTGGGCTCAAGAGAACAGCTACTCTTGCTAAA 1260
Db 395 rAlaIleArgSerAspGlyProHisProGlyArgArgLeuAspSerAlaLeuSerGluSe 415
Qy 1261 ACCA-----TCACTGTGTGAGGTACACCCAGCTACCC-----TTCCGAAGCCA 1302
Db 415 rProLysThrAlaLeuLeuAlaLeuSerArgProArgProProIleThrArgSerGl 435
Qy 1303 AAGTATGATCAGGAGGTGGACCTGAGCTGAGCTGAGCCGAGATGACCTGGATGGAGGCAG 1362
Db 435 nSerTyrIleThrLeuTrp-----GluGluLy 444
Qy 1363 GAGGAGCAGGAGGAGCAGGAGATTTCGGGAAAT 1393
Db 444 sGlnGlnArgLysLysGluLysSerGlySer 454

RESULT 10
Q921G9
ID Q921G9 PRELIMINARY; PRT; 904 AA.
AC Q921G9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to tight junction protein 3.
GN TJP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012518; AAI12518.1;
DR MGD; MGI:1351650; TJP3.
DR InterPro; IPR00619; Guanylate_kin.
DR Pfam; PF00625; Guanylate_kin; 1.
DR PRINTS; PF00595; PDZ; 3.
DR PRINTS; PRO1597; ZONOCCLUDNS.
DR PROSITE; PS01600; ZONOCCLUDNS3.
DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 904 AA; 99169 MW; A4D9C58AEEC485CC CRC64;

Alignment Scores:
Pred. No.: 1.7e-10 Length: 904
Score: 275.00 Matches: 134
Percent Similarity: 34.51% Conservative: 90
Best Local Similarity: 20.65% Mismatches: 191
Query Match: 6.67% Indels: 234
DB: 11 Gaps: 24

US-09-502-945-4 (1-2236) x Q921G9 (1-904)
Qy 385 GGCCTCGCCTGAGTGTGCTGGTGGCTGGAGTTGGCTGTGGG---CTCTTCATCTCC 441
Db 21 GlyPheGlyIleAlaValSerGlyGlyHisAspArgAlaSerGlySerValValSer 40
Qy 442 CACCTCATCAAGCGGTCAGCAGCAGCAGCGTCCGGTCCAGGTAGGGAGAGATCGTC 501
Db 41 AspValValProGlyGlyProAlaGluGly---ArgLeuArgThrGlyAspHisIleVal 59
Qy 502 CGGATCATGGATATTCATCTCTCTGATCCCATGAGGAGGTCATCAACCTCATTCGA 561
Db 60 MetValAsnGlyValSerValGluAsnValThrSerAlaPheAlaIleGlnIleLeuLys 79
Qy 562 ACC---AAGAAACTGTGTCTCATCAAAAGTG-----AGACACATCGCGCTGATCCCC 609
Db 80 ThrCysThrLysThrAlaAsnValThrValLysArgProArgValGlnLeuProAla 99

```


Score:	272.00	Matches:	69
Percent Similarity:	52.74%	Conservative:	37
Best Local Similarity:	34.33%	Mismatches:	83
Query Match:	6.60%	Indels:	12
DB:	4	Gaps:	5

US-09-502-945-4 (1-2236) x Q9NTP3 (1-202)

QY	379	CCCGAAGGC---CTCGGCCGTGAGTGGCTGGTGTCGCTGAGTTTGGCTGTGGGCTTC	433
Db	6	ProAlaGlyArgLeuGlyPheSerValArgGlyGlySerGluHisGlyLeuGlyIlePhe	25
QY	436	ATCTCCACCATCAAAAGCGGTCCAGGCAGACAGCGTGGGCTCCAGTAGTGGGACGAG	495
Db	26	ValSerLysValGlutGluGlySerSerAlaGluArgAlaGlyLeuCysValGlyAspLys	45
QY	496	ATGCTCCGGATCAATGAGATTCCATCTCCTGTACCCTGAGGAGTCAATCAACTC	555
Db	46	IleThrGluValAsnGlyLeuSerLeuGluSerThrThrMetGlySerAlaValLysVal	65
QY	556	ATTGGAACCAAGAANAAGTGTCTCATCAAGTAGCAGACATCGGCCTGATCCCCGTGAAA	615
Db	66	LeuThrSerSerSerArgLeuHisMetMetValArgMetGlyArgValProGlyIle	85
QY	616	AGCTCTCCTGATGAGCCCTCACCTTGGCAGTATGTGGAT-----CAGTTTGTG	663
Db	86	LysPheSerLysGluLysThrTrp----ValAspValValAsnArgargLeuVal	103
QY	664	TCGGAACTCGGGCGTGCAGGACGCTGGCTGCCCTGGAAATCGGGAACAAGGAG	723
Db	104	ValGluLys-----CysGlySerThrProSerAspThrSerSerGluAspGlyVal	120
QY	724	AAGAAGTCTTTCATCAGCCGTGAGTCCCAGGC-----CTTGGCTGCAGCATTTCC	777
Db	121	ArgarginIleValHisLeuTyrrThrThrSerAspSpPheCysLeuGlyPheAsnIleArg	140
QY	778	AGCGGCCCCATCCAGAAGCCTGGCATCTTTATCAGCCATGTGAAACCTGGCTCCGTCT	837
Db	141	GlycylLysGluPheGlyLeuGlyIleTyrrValSerLysValAspHisGlyLeuAla	160
QY	838	GCTGAGTGGGATTTGAGATAGGGACACAGATTGTCCAAGTCAATGGCTCGACTTCTCT	897
Db	161	GluGluAsnGlyIleLysValGlyAspGlnValLeuAlaAlaAsnGlyValArgPheAsp	180
QY	898	AACTGTGATCACAGAAGGCGTAAATGTGCTGAAAAATAGCCGACGCTGACCATCC	957
Db	181	AspIleSerHisSerGlnAlaValGluValLeuLysGlyGlnThrHisIleMetLeuThr	200
QY	958	ATT 960	
Db	201	Ile 201	

RESULT 12

ID	075970	PRELIMINARY;	PRT;	2042 AA.
AC	01-NOV-1998 (TREMBlRel. 08, Created)			
DT	01-NOV-1998 (TREMBlRel. 08, Last sequence update)			
DT	01-DEC-2001 (TREMBlRel. 19, Last annotation update)			
DE	Multi PDZ domain protein MUPPI.			
DE	MUPPI.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RA	Eng L., Kravitsky G., Clapham D.E.;			
RT	"Human homolog of MUPPI protein.";			
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF093419; AAC61870.1; -			
DR	HSSP; Q12959; 1PDR.			

QY 701 CTGGAAATCGGAAACAAAGGAGAGAGGCTTTCATCAGCGCTGGTAGCTCCCGAGGCC 760
Db 1830 InGlyLeuArgThrValGluMetLysLys-----GlyProThrAspSerL 1845
QY 761 TTGGCTGAGCATTTCCAGCGC-----CCCATCCAGAGCCTGGCATCTTTATCA 811
Db 1845 euGlySerIleAlaGlyValGlySerProLeuGlyAspValProIlePheIleA 1865
QY 812 GCATGTGCAACCTGGCTCCCTGCTGCTGAGGTG---GGATTGGAGATAGGGACACCA 868
Db 1865 IamMetHisProThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgI 1885
QY 869 TTGTCCAAAGTCAATGGCTCGACTTCTTAACCTGGATCACAAGAGAGCTGTAAATGTC 928
Db 1885 leValThrIleCysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValAsnLeuL 1905
QY 929 TGAATAATAGCGCGACCTGACCATCTCCATTGTAGCTGCAGCTGGCGGGAGCTGTTC 988
Db 1905 euLysAsnAla---SerGlySerIleGluMetGlnValAlaGlyGlyAspVal--- 1922
QY 989 TCAGACACGGGAGCGGTGGTGGAGGCGCGCAGCGTGCAGCTGCAGCGGAGGAGCTC 1048
Db 1923 -----SerValValThrGlyHisGlnGlnGluProAlaSerSerLeus 1938
QY 1049 TCATCAGAGCGGTGGCTGGAGTCCACAAGATCTCCAGGAGCAGGAGAGATGG 1108
Db 1938 exPhe-----ThrGlyLeuThrSerSerIlePheGln----- 1949
QY 1109 AGCGGCAAGAGAGAAATGGCCAGAGGAGCAGAGGAAATGAGAGATACCGGA 1168
Db 1949 ----- 1949
QY 1169 AGGAGATGAACAGATTTGTAGAGGAGAGAGAGTTTAAAGCAATGGGAAGAGACT 1228
Db 1950 -----AspAspL 1952
QY 1229 GGGGCTCAAGGAACAGTACTTCTTGCCTTAAACATCATCTGCTGAGGTACACCCAGTAC 1288
Db 1952 euGlyProGln-----CysLysSerIleThrLeuGluArgGlyPro--- 1966
QY 1289 CCCTTCGCAAGCAAAGTATGATGATCAGGGAGTGGAACTCAGCTCGAGCCCGCAGATGACC 1348
Db 1966 ----- 1966
QY 1349 TGGATGAGCGCAGGAGGAGCAGGAGAGAGGATTCGGAATATGAGGAAGGCTTGT 1408
Db 1967 --AspGly----- 1968
QY 1409 ACCCTACTCTATGTTACCCCGCAGCAGATCATGGGGAGGATGTCGGGCTCCTACGCA 1468
Db 1968 ----- 1968
QY 1469 TCAAGAAGGAGGATCCTTACAGCTGGCCCTGGAAGCGGTGTGGACTCCCGCATGGG- 1527
Db 1969 -----LeuGlyPheSerIleValGlyGlyThrGlySerProHisGlyA 1983
QY 1528 -----AAGTGTGCTGTTCTGCTGATCAGCGGGAGCTGCTGAGCGGCATGTGGCA 1582
Db 1983 spLeuProIleThrValPheAlaLysGlyAlaAlaSerGluAspGlyArgL 2003
QY 1583 TTGTGAAGGGCAGAGATCATGGCAATCAACGGCAAGATGTGACAGACTACACCTCG 1642
Db 2003 euLysArgGlyAspGlnIleIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisG 2023
QY 1643 CTGAGCTGACGCTGCCCTGCAGAG 1668
Db 2023 luGluAlaValAlaIleLeuLysArg 2031
RESULT 13
O55164
ID O55164 PRELIMINARY; PRT; 2054 AA.
AC O55164;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Multi PDZ domain protein 1.
GN MUPP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
OX [1]
RN SEQUENCE FROM N.A.
RN RN
TI TISSUE=BRAIN;
RA MEDLINE=98136865; PubMed=9537516;
RT "Cloning and characterization of MUPP1, a novel PDZ domain protein.";
RL FEBS Lett. 424:63-68(1998).
DR EMBL; AJ001320; CAA04681.1; .
DR HSP; Q12959; IPRD.
DR InterPro; IPR004172; L27.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF02828; L27; 1.
DR Pfam; PF00595; PDZ; 13.
DR SMART; SM00228; PDZ; 13.
DR PROSITE; PS01016; PDZ; 13.
SQ SEQUENCE 2054 AA; 218591 MW; 44BD3F42B801F78F CRC64;

Alignment Scores:
Pred. No.: 6,13e-10 Length: 2054
Score: 269.00 Matches: 120
Percent Similarity: 36.64% Conservative: 72
Best Local Similarity: 22.90% Mismatches: 160
Query Match: 6.53% Indels: 172
DB: 17

US-09-502-945-4 (1-2236) x O55164 (1-2054)
QY 157 GCAGAGAGGACTATCTATGATGCTGCGAATGATACACAGACCATGAGCTGSGCC 216
Db 1672 AlaThrHisAspGluAlaIleAsnValLeuArgGlnThrProGlnArgVal----- 1688
QY 217 GTGTCGTGGGAGAGCCTGAAGCTGGTCTATCAATGAACCCAGCCGCTCTGCTCTGTTGAT 276
Db 1689 -----ArgLeuThrLeuTyr--- 1693
QY 277 GCCATTCGCGCTGATCCACTGAAGCACCAGGTGGATATGATCAGCTACCCCGCG 336
Db 1694 -----ArgAspGluAlaProTyrLysGluArgValCysAspThrPheThr----- 1709
QY 337 CGCTCCAGGAAGCTGAAGGAGGTGCTGACCGTCTGCACCCGCGAGGCTCGGCTG 396
Db 1710 -----ValGluLeuGlnLysArgProGlyLysGlyLeuGlyLeu 1722
QY 397 AGTGTGCTGGTGGCTGGAGTTTGGCTGTGGCTCTTCATCTCCCATCATCAAGGC 456
Db 1723 SerIleValGlyLysArgAsn---AspThrGlyValPheValSerAspIleValLysGly 1741
QY 457 GGTCCAGGACAGCGCTCGG---CTCCAGGTAGGGAGAGATCGTCCGAGATCATGGA 513
Db 1742 GlyIleAlaAspAlaAspGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGly 1761
QY 514 TATTCATCTCTCTCTGATCCCATGAGGAGGTCTATCAACCTCATTCGA---ACCAAGAA 570
Db 1762 GluAspValArgAsnAlaThrGlnGluAlaValAlaLeuLeuLysCysSerLeuGly 1781
QY 571 ACTGTGTCATCAAAAGTGAGACATCGGCTGATCCCGGTGAAAGATCTCTCTGATGAG 630
Db 1782 ThrValThrLeuGluValGlyArgIleLysAlaAlaProPheHisSerGluArgArgPro 1801
QY 631 CCCCTCACTGGCAGTATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 690
Db 1802 SerGlnSerSerGlnValSerLeuSerLeuSerLeuSerLeuSerLeuSerLeuSer 1821
QY 691 CTGGGCTCCCTGGAAATCGGGAACCAAGAGAGAGAG----- 729


```
Db 1822 IieHisThrSerGluSerSerGluSerSerAlaLysLysAsnAlaLeuAlaSerGluIle 1841
QY 730 -----GTCATCATCAGCCCTGGTAGGCTCCCGAGGCTTGGCTGCAGCATT 774
Db 1842 GlnGlyLeuArgThrValGluIleLysLysGlyProAlaAspAlaLeuGlyLeuSerIle 1861
QY 775 TCCAGCGGC-----CCATCCAGAGCCTGGCACTTTATCAGCCATGTGAACCT 825
Db 1862 AlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetHisPro 1881
QY 826 GCCTCCCTGCTGCTGAGGTG---GATTGGAGATAGGGGACGACATTCGCAAGTCAAT 882
Db 1882 AsnGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleCys 1901
QY 883 GCGCTGCACTTCTTAACCTGGATCACAAGAGGCTGTAAATGTCTGAAAAATAGCCGC 942
Db 1902 GlyThrSerAspGlyMetThrHisThrGlnAlaValAsnLeuMetLysAsnAla--- 1920
QY 943 AGCCTGACCATCTCCATTGCTAGCTGCAGCTGGCCGGGAGCTGTTATGACACACCGGAG 1002
Db 1921 SerGlySerIleGluValGlnValAlaGlyGlyAspVal----- 1934
QY 1003 CGCTGCGCAGAGCGCGCAGCTGAGCTGCAGCGCAGGAGCTTCTCATGCAGAAGCGG 1062
Db 1935 SerValValThrGlyHisGlnGlnGluLeuAlaAsnProCysLeuAlaPhe----- 1951
QY 1063 CTGGCGATGGATCCAAACAGATCTCCAGGAGCAGCAGGATGGAGCGCAAGAGAGA 1122
Db 1952 ThrGlyLeuThrSerSerThrIle----- 1959
QY 1123 AAGAAATGCCCCAGAGCAGCAGGAGGAAATGAGAGATACCGGAGGAGTGAACAG 1182
Db 1959 ----- 1959
QY 1183 ATTGTAGAGGAGGAGAGAGAGTTTAAGAGCAATGGAGAGAGACTGGGGTCAAGGAA 1242
Db 1960 -----PheProAspLeuGlyProProGln 1968
QY 1243 CAGCTACTCTGCCATAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCA 1302
Db 1969 Ser-----LysThrIleThr----- 1973
QY 1303 AAGTATGATCAGGAGTGAACCTGAGCTGAGCGCGCAGAGTACGCTGGATGGAGGCAGC 1362
Db 1974 -----LeuAspArgGlyPro 1978
QY 1363 GAGGAGCAGGAGAGCAGGATTTCGGAATATGAGAGAGGCTTTGACCCCTACTCTATG 1422
Db 1979 Asp----- 1979
QY 1423 TTCACCCAGACAGCAGATCATNGGGAAGGATGTCGGCTCTTACGATCAAGAAGAGGGA 1482
Db 1979 ----- 1979
QY 1483 TCCTTAGACCTGGCCCTGGAAGCGGTGGGACTCCCCCATTTGGG-----AAGGTGTC 1536
Db 1980 GlyLeuGlyPheSerIleValGlyGlyrGlySerProHisGlyAspLeuProIleTyr 1999
QY 1537 GTTCTGCTGTATGAGCGGGAGCTGCTGAGCGGCATGTTGGCATTCGTGAAGGGGAC 1596
Db 2000 ValLysThrValPheAlaLysGlyAlaAlaGluAspGlyArgLysArgGlyAsp 2019
QY 1597 GAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCCCTGGCTGAGCTGACGCT 1656
Db 2020 GlnIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaValAla 2039
QY 1657 GCCTGTCAGAG 1668
Db 2040 IleLeuLysArg 2043
RESULT 14
Q96C69
```

```
ID Q96C69 PRELIMINARY; PRT; 835 AA.
AC Q96C69;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 87.7 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014632; AAH14632.1; -.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR PROSITE; PS50106; PDZ; 3.
KW Hypothetical protein.
FT NON_TER 1
FT SEQUENCE 835 AA; 87669 MW; F59DF419D7AAF991 CRC64;
SQ SEQUENCE 835 AA; 87669 MW; F59DF419D7AAF991 CRC64;

Alignment Scores:
Pred. No.: 8,76e-10 Length: 835
Score: 264.50 Matches: 190
Percent Similarity: 32.75% Conservative: 93
Best Local Similarity: 21.99% Mismatches: 233
Query Match: 6.42% Indels: 350
DB: 4 Gaps: 36

US-09-502-945-4 (1-2236) x Q96C69 (1-835)
QY 289 CTGATCCCACTGAAGCAGCAGGTGGTAAATATGATCAGCTGACCCCGCGCTCCAGGAAG 348
Db 30 IleThrProLeuArgProGluAspAspTyrSerPro-----ArgGluArg 45
QY 349 CTGAAGGAGGTGCTGCTGGACCGCTGTCGCCGCCGAA----- 384
Db 46 GlyGlyGlyLeuArgLeuProLeuProGlySerProGlyProLeuArgGlnArg 65
QY 385 -----GGCCTCGGCTGAGTGTGCGTGGC----- 411
Db 66 HisValAlaCysLeuAlaArgSerGluArgGlyLeuGlyPheSerIleAlaGlyLys 85
QY 412 -----CTGGAGTTTGGCTGCTGCTTCATCTCCACCTCATCAAGGC 456
Db 86 GlySerThrProTyrArgAlaGlyAspAlaGlyIlePheValSerArgIleAlaGly 105
QY 457 GGTGAGGAGCAGCGTCCGG-----CTCAGTAGGGAGAGATGCTCCGGATCAATGGA 513
Db 106 GlyAlaAlaHisArgAlaGlyThrLeuGlnValGlyAspArgValLeuSerIleAsnGly 125
QY 514 TATTCATCTCTCTGCTGACCATGAGGAGTGCATCACTCATT---CGAACCAAGAA 570
Db 126 ValAspValThrGluAlaArgHisAspHisAlaValSerLeuLeuThrAlaAlaSerPro 145
QY 571 ACTGTGTCATCAAAAGTGAGACACATCGCCTGATCCCGTGAAAGCTCTCCT----- 624
Db 146 ThrIleAlaLeuLeuLeuGluArgGluAlaGlyGlyProLeuProProSerProLeuPro 165
QY 625 -----GATGACCCCTCACTTGGCAGTATGTCAGTATGTCGGAATCTGGGGC 678
Db 166 HisSerSerProThrAlaAlaValAlaThrSerIleThrThrAlaThrProGly 185
QY 679 GTGCGAGC----- 687
Db 186 ValProGlyLeuProSerLeuAlaProSerLeuAlaAlaLeuGluGlyProTyr 205
QY 688 -----AGCCTGGCTCCCTCGGAATCGGAAACCAAGAG 723
Db 206 ProValGluGluIleArgLeuProArgAlaGlyPro----- 218
```


OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
EX	MEDLINE=96127530; PubMed=8590280;
RA	Nagase T., Seki N., Tanaka A., Ishikawa K., Nomura N.;
RT	"Prediction of the coding sequences of unidentified human genes. IV.
RT	"The coding sequences of 40 new genes (K1AA0121-K1AA0160) deduced by
RT	analysis of cDNA clones from human cell line KG-1.";
RL	DNA Res. 2:167-174(1995).
DR	EMBL; D63481; BAA09768.2; -.
DR	HSSP; Q12923; 3PDZ.
DR	InterPro; IPR001611; LRR.
DR	InterPro; IPR003591; LRR_typ.
DR	InterPro; IPR001478; PDZ.
DR	Pfam; PF00360; LRR; 11.
DR	Pfam; PF00595; PDZ; 4.
DR	SMART; SM00369; LRR_TYP; 9.
DR	SMART; SM00228; PDZ; 4.
DR	PROSITE; PS50106; PDZ; 4.
FT	NON_FER 1
FT	NON_FER 1
SQ	SEQUENCE 1564 AA; 166998 MW; 07E1A44CF805BE56 CRC64;

Alignment Scores:					
Pred. No.:	1-32e-09	Length:	1564		
Score:	263.50	Matches:	190		
Percent Similarity:	32.75%	Conservative:	93		
Best Local Similarity:	21.99%	Mismatches:	233		
Query Match:	6.40%	Indels:	350		
DB:	4	Gaps:	36		
US-09-502-945-4 (1-2236) x Q14160 (1-1564)					
QY	289	CTGATCCCACTGAAGCACCAGGTGGGATATGATCATCGTGCACCCCGGCGCTCCAGGAAG	348		
		::: ::: ::: :::			
Db	759	IleThrProLeuArgProGluAspSerPro-----ArgGluAraArg	774		
QY	349	CTGAAGGAGGTGGGTCTGGACCTCTGCACCCGAA-	384		
		:::			
Db	775	GlyGlyLeuArgLeuProLeuProGluSerProGlyProLeuArgGlnArg	794		
QY	385	-----GGCTCGGCGTGAAGTGCTGGTGGTGC---	411		
		::: :::			
Db	795	HisValAlaCysLeuAlaArgSerGluArgGlyLeuGlyPheSerIleAlaGlyGlyLys	814		
QY	412	-----CTGGAGTTTGGCTGTGGGCTCTTCATCTCCCACCTCATCAAGGC	456		
		: : ::: :::			
Db	815	GlySerThrProTyrrArgAlaGlyAspAlaGlyIlePheValSerArgIleAlaGluGly	834		
QY	457	GGTCAGGCAGACAGCGTCGGG---CTCCAGGTAGGGGACGAGATCTCCGGATCAATGA	513		
Db	835	GlyAlaAlaHisArgAlaGlyThrLeuGlnValGlyAspArgValLeuSerIleAsnGly	854		
QY	514	TATTCCATCTCTCTGTATCCCATAGGAGGTGTCATCAACTCATT---CGAACCAAGAAA	570		
		: : : : : : : : : : : : : : :			
Db	855	ValAspValThrGluAlaArgHisAspHisAlaValSerLeuLeuThrAlaAlaSerPro	874		
QY	571	ACTGTGTCCATCAAGTGACACATCGCCGTGATCCCCGTGAAAGCTCTCCT---	624		
		: : : : : :			
Db	875	ThrIleAlaLeuLeuGluArgGluAlaGlyGlyProLeuProSerProLeuPro	894		
QY	625	-----GATGAGCCCTCATCTTGGCAGTAGTGTGGATCAGTTTTGTGCGGAANTCGGGGC	678		
		: : : : :			
Db	895	HisSerSerProThrAlaAlaValAlaThrThrSerIleThrThrAlaThrProGly	914		
QY	679	GTCCGAGGC-----	687		
Db	915	ValProGlyLeuProSerLeuAlaProSerLeuLeuAlaAlaLeuGluGlyProTyrr	934		
QY	688	-----ACGCTGGGCTCCCTCGGAANTCGGGAACAANGAG	723		
Db	935	ProValGluLeuIleArgLeuProArgAlaGlyGlyPro-----	947		


```
OY 1410 C-----CCCTACTCTATGTTACCCAGAGCAGATC-----ATGGGGAA 1448
Db 1296 uLeuGluValArgValProGlnAlaGluGlyProProlsArgValSerLeuValGlyAl 1316
OY 1449 GGAT-----GTCCGGCTCTACCATCACAAGAGGAGG 1481
Db 1316 aAspAspLeuArgLysMetGlnGluGluAlaArgLysLeuGlnLysArgAlaG1 1336
OY 1482 ATCCCTTAGACCTGGCCCTGAAGGGGGTGTGGACTCCCCCATTTGGGAAGGTGTCGTTTC 1541
Db 1336 nMetLeuArgGluAlaAlaGluAlaGlyAlaGluAlaArg----- 1349
OY 1542 TGCTGTGTATGACGGGGGAGCTGCTGAGCGCATGGTGGCATTTGTCAAAGGGGACGAGAT 1601
Db 1349 ----- 1349
OY 1602 CATGGCAATCAAGGCAAGATTGTGACAGACTACACCTGGCTGAGGTGACGCTGCCCT 1661
Db 1350 -LeuAlaLeuAspGlyGlu-----ThrLeuGlyGlu----- 1359
OY 1662 GCAGAAGGCTGGAATCAGGGGGGACTGGATCGACCTTGTGTTGCGTCTGCCCCOC 1721
Db 1359 ----- 1359
OY 1722 AAAGGAGTATGACGATGACGTACCTTCTGTGCTGAAGTCCAAAGGGGAACCAATTCA 1781
Db 1360 -GluGluGlnGluAspGluGln-ProProTrpAlaSerPro-Ser-----ProThrSer 1376
OY 1782 CGCGTTAGGAACAGTGACTCG-----GCCCCACCTCGTGNACACAAA 1826
Db 1377 ArgGlnSerProAlaSerProProProLeuGlyGlyGlyAlaProValArgThrAlaLys 1396
OY 1827 GCCTCGGACGACCTTGAGAGAGGCCACATGACACACACAGATGGCATCCTTGGGACCT 1886
Db 1397 Ala-----GluArgArgHis----- 1401
OY 1887 GAATCTATCACCAGGAATCTCAAACCTTGGCCCTGAACAGGCGCCAGATAAGGAA 1946
Db 1402 -----GlnGluArgLeuArgValGlnSerProGluProProAlaProGluArg 1417
OY 1947 CAGCTGGGGCCACTTTT-TTGAAGGCCAATGTGGAGGAAGGGAGCAGCCGCTTTGG 2005
Db 1418 AlaLeuSerProAlaGluLeuArgAla-----LeuGluAlaGluLysArgAlaLeuTrp 1435
OY 2006 GAGAAGATCTCAAGGATCCAGACTCTCATTCCTTCTCTGCGCCAGTGAATTTGGTCTC 2065
Db 1436 ---ArgAlaAlaArgMetLysSerLeuGluGlnAspAlaLeuArgAlaGlnMetValLeu 1454
OY 2066 TCCAGCTTTGG-----GGGACT----- 2083
Db 1455 SerArgSerGlnGluGlyArgGlyThrArgGlyProLeuGluArgLeuAlaGluAlaPro 1474
OY 2084 ---CCTTCCTTGACCTTAATAGACCCGACTGGAGTCTCTC----- 2122
Db 1475 SerProAlaProThrProSerProThrProValGluAspLeuGlyProGlnThrSerThr 1494
OY 2123 -----TCTCTCCATCCCTCTCCT 2140
Db 1495 SerProGlyArgLeuSerProAspPheAlaGluLeuArgSerLeuGluProSerPro 1514
OY 2141 CTGCC 2146
Db 1515 SerPro 1516
```

Search completed: March 21, 2003, 12:59:28
Job time : 108.681 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 08:38:27 ; Search time 52.8072 Seconds
(without alignments)
10910.938 Million cell updates/sec

Title: US-09-502-945-5

Perfect score: 3989

Sequence: 1 cctggccggtgcggtgcg.....tatttcccgcttaaaaaa 2162

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPro.spool/US09502945/runat_14032003_101057_19087/app_query.fasta_1.10979
-DB=A_Geneseq_101002 -QPMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptio -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09502945_@CGN_1_1_365 @runat_14032003_101057_19087 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	3196.5	80.1	652	20	AA107094
2	3196.5	80.1	652	20	AA107050
3	2188	54.9	481	20	AA107098
4	2168	54.3	521	20	AA107099
5	2046	51.3	403	20	AA107095
6	1557.5	39.0	328	20	AA107097
7	1175.5	29.5	268	20	AA138440
8	435	10.9	120	21	AA153238
9	418	10.5	493	22	ABB61924
10	417	10.5	98	20	AA138390
11	406	10.2	95	20	AA138437
12	271	6.8	1535	23	AA1087934
13	267.5	6.7	1037	22	ASG22366
14	267	6.7	928	23	AA121718
15	263	6.6	856	21	AA101383
16	256	6.4	2037	21	AA153753
17	256	6.4	763	20	AA104741
18	256	6.4	1005	20	AA104731
19	256	6.4	1373	20	AA104730
20	256	6.4	2000	20	AA104732
21	256	6.4	2070	20	AA104733
22	254	6.4	1111	23	AA1087918
23	250.5	6.3	1881	20	AA124025
24	249	6.2	1552	22	AA108793
25	249	6.2	1554	22	ASG07288
26	237	5.9	46	22	AB138777
27	237	5.9	46	22	AB138777
28	237	5.9	46	22	AA159416
29	237	5.9	46	22	AA11974
30	237	5.9	46	22	AA119386
31	237	5.9	46	22	AA132249
32	237	5.9	46	23	ASG41789
33	236	5.9	1745	23	AB157253
34	234.5	5.9	767	23	AA1084269
35	234.5	5.9	767	23	AA1084328
36	234	5.9	1526	22	AA109777
37	232.5	5.8	767	21	AA122138
38	232.5	5.8	817	19	AA148101
39	232.5	5.8	817	20	AA130137
40	232.5	5.8	849	19	AA148102
41	232	5.8	45	22	AB138778
42	232	5.8	45	22	AB138778
43	232	5.8	45	22	AA159417
44	232	5.8	45	22	AA171975
45	232	5.8	45	22	AA119387

ALIGNMENTS

RESULT 1
AA107094
ID AA107094 standard; Protein; 652 AA.

XX AC AA107094;

XX DT 02-JUL-1999 (first entry)

XX DE Colon cancer associated antigen precursor sequence.

XX KW Cancer associated antigen; diagnosis; research; treatment; human;

XX KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;

XX KW prostate cancer.

XX OS Homo sapiens.

XX PN WO9904265-A2.

XX XX 28-JAN-1999.

PD

XX PF 15-JUL-1998; 98WO-US14679.
 XX PR 22-JUN-1998; 98US-0102322.
 XX PR 17-JUL-1997; 97US-0896164.
 XX PR 10-OCT-1997; 97US-0061599.
 XX PR 10-OCT-1997; 97US-0061765.
 XX PR 10-OCT-1997; 97US-0948705.
 XX PR 11-OCT-1997; 97GB-0021697.
 XX PA (LUDW-) LUDWIG INST CANCER RES.
 XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 XX WI; 1999-132448/11.
 XX PT New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 XX PS Disclosure; Page 656-658; 787pp; English.
 XX CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 XX SQ Sequence 652 AA;
 Alignment Scores:
 Pred. No.: 2,77e-268 Length: 652
 Score: 3196.50 Matches: 626
 Percent Similarity: 96.02% Conservative: 1
 Best Local Similarity: 95.87% Mismatches: 0
 Query Match: 80.13% Indels: 26
 DB: 20 Gaps: 1
 US-09-502-945-5 (1-2162) x AAY07094 (1-652)
 QY 97 ATGGACCAAGAGTGGCCGAGAAATCCGGCATAGGTGGATTTCTGATTGAAAATGAT 156
 Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
 QY 157 CGAGAGAAGGACTATCTCTATGATGTGCTGGGAATGTACCACACAGACCATGGAGCTGGCC 216
 Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
 QY 217 GTGCTGTGGAGACCTGAAGCTGGTCAATCAATGAACCCAGCGCTCGCTCTGTTGAT 276
 Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
 QY 277 GCATTGGCCGCTGATCCCACTGAAGCACAGGTGGAAATATGATCAGCTGACCCCGGG 336
 Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 QY 337 CGCTCCAGGAAGCTGAAGAGGTGGCTGTGGACCGCTCTGCACCCCGAAGGCGCTGGGCTG 396
 Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeu 100
 QY 397 AGTGTGCGTGGGCGCTCGAGTTTGGCTGTGGGCTCTTCATCTCCCACTCATCAAGGC 456
 Db 1

Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
 QY 457 GGTTCAGGCAGACAGCGTCCGGCTCCAGGTAGGGACGAGATCGTCCGGATCAATGATAT 516
 Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 QY 517 TCATCTCTCTCTCTACCCATGAGGAGGTCAACACCTCATTCGACCAAGAAACTGTG 576
 Db 141 SerIleSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
 QY 577 TCATCAAAAGTGAGACACATCGCGCTGATCCCGTGAAGCTCTCTCATGAGCCCTC 636
 Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
 QY 637 ACTTGGCAGTATGTGATCAGTTGTGTCGGAATCTGGGGCGCTGCGAGCGACCTGGGC 696
 Db 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlySerLeuGly 200
 QY 697 TCCCTGGAAATCGGGAACAAACAGGAGAGAGAGTCTTTCATGAGCTGTAGCTCCCA 756
 Db 201 SerProGlyAsnArgGluAsnLysGluLysValPheIleSerLeuValGlySerArg 220
 QY 757 GGCCTTGGCTGAGACATTTCCAGCGCCCATCCAGAACCTGGCATCTTTATCAGCCAT 816
 Db 221 GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
 QY 817 GTGAAACCTGGTCCCTGTCTGTGAGGTGGATTTGAGATAGGGACACAGATTGTGAA 876
 Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
 QY 877 GTCATATGGGCTGACCTCTCTAACCTGGATACAAAGAGGCTGTAATGTGCTGAAAT 936
 Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
 QY 937 AGCCGAGCGCTGACCATCTCCATTGTGAGTGCAGTGGCGGGAGCTGTTTCATGACAGAC 996
 Db 281 SerArgSerLeuThrIleSerIleValAlaAlaAlaGlyArgGluLeuPheMetThrAsp 300
 QY 997 CGGAGCGGCTGGCAGAGCGCGGCGAGCTGAGCTCAGCGGAGGAGTCTTCATGCGAG 1056
 Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
 QY 1057 AACCGGCTGGCGATGGAGTCCAAACAGATCTCCAGGAGCAGGAGATGGAGCGCAA 1116
 Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGluGlnGluMetGluArgGln 340
 QY 1117 AGGAGAAAAGAAATTCGCCAAGAGGAGGAGAGAAATAGAGATACCGAAGAGAGATG 1176
 Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
 QY 1177 GAACAGATTGTAGAGGAGGAAGAAGTTTAAAGAACAATGGGAAGAAGCTGGGGCTCA 1236
 Db 361 GluGlnIleValGluGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380
 QY 1237 AAGAACAGCTACTCTTTCCTTAAACCATCAGTCTGCTAGGATACCCAGTACCCCTTCG 1296
 Db 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 QY 1297 AAGCCA----- 1302
 Db 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
 QY 1303 -----AAGTATTTCCGGAATATGAGGAAGGCTTTCACCCCTA 1341
 Db 421 GlyThrGluGluGlnGlyGluGln-AspPheArgLysTyrGluGluGlyPheAspProTy 440
 QY 1342 CTCTATGTTTCAACCCAGCAGATCATGGGAAGGAGTCCCGCTCTCTAGCATCAGAA 1401
 Db 440 rSerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLy 460
 QY 1402 GGAGGGATCTTAGACCTGGCGCTGGAGCGGTGTGGACTCCCTCCCATTCGGAGGTGT 1461
 Db 460 sGluGlySerLeuAspLeuAlaLeuGluGlyGlyValAspSerProIleGlyLysValva 480

QY 1462 CGTTTCTGCTGTATGAGCGGGAGCTGCTGAGCGGCATGGTGGCATTTGTCARAGGGGA 1521
Db 480 lvalserAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAs 500
QY 1522 CGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTCGCTGAGGCTGACGC 1581
Db 500 pglulleMetaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAl 520
QY 1582 TGCCCTCGAGAGGCTGGAATCAGGGGGGAGCTGGATCGACCTGTGGTTCGCGCTG 1641
Db 520 aAlaLeuGlnLysAlaTyrAsnGlnGlyAspTyrIleAspLeuValValAlaValCy 540
QY 1642 CCCCCCAAGAGTATGAGTACGAGCTGACCTTCTGCTGAAGTCCAAAGGGGAACCA 1701
Db 540 sProProLysGluTyrAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnG 560
QY 1702 AATTCACCGCTTAGGAACAGTACGCTCGCGCCACCTCGTGAACACAAAGCCCTCGGAC 1761
Db 560 nIleHisAlaLeuGlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgTh 580
QY 1762 CAGCCTTGAGAGGCCACATGACACACACAGATGGCATCCTTGGGACCTGAATCTATC 1821
Db 580 rSerLeuGluArgGlyHisMetThrHisThrArgTyrHisProTyrAspLeuAsnLeuSe 600
QY 1822 ACCCAGGAATCTCAAACTCCCTTTGCGCCCTGAACAGCGCCAGATAAGACAGCTCGGG 1881
Db 600 rProArgAsnLeuLysLeuProLeuAlaLeuAsnGlnGlyGlnIleArgAsnSerSerG 620
QY 1882 CCACCTTTTTTGAAGGCCAATGTGGAGGAAAGGAGGAGCAGCCGCTTTGGGAGAAGATCT 1941
Db 620 yHisPheGluGlnGlyCysGlyGlyLysGlyAlaAlaSerArgLeuGlyGluAspLe 640
QY 1942 CAAGGATCCAGACTCTATCTCTTCTTCTGCGCCAG 1978
Db 640 ulysaspProaspSerHisSerPheProLeuAlaGln 652
RESULT 2
AAY07050
ID AAY07050 standard; Protein: 652 AA.
AC AAY07050;
XX
XX 02-JUL-1999 (first entry)
DE Renal cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX Homo sapiens.
XX
XX W09904265-A2.
XX
XX 28-JAN-1999.
XX
XX 15-JUL-1998; 98WO-US14679.
XX
XX 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI: 1999-132448/11.

XX

PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers

XX Disclosure; Page 436-438; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

XX Sequence 652 AA;

Alignment Scores:
Pred. No.: 2,77e-268 Length: 652
Score: 3196.50 Matches: 626
Percent Similarity: 96.02% Conservative: 1
Best Local Similarity: 95.87% Mismatches: 0
Query Match: 80.13% Indels: 26
DB: 20 Gaps: 1

US-09-502-945-5 (1-2162) x AAY07050 (1-652)

QY 97 ATGACCGCAAAAGTGGCCGAGAAATCCGCGCATAGGTGGATTCTTCTGATTGAAATGAT 156

Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20

QY 157 GCAGAGAGGACTATCTCTATGATGTCTGCGAATGTACCACAGACCATGGAGCTGGCC 216

Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40

QY 217 GTGCTCGTGGGAGACCTGAAGCTGTCATCAATGAACCCAGCGCTCTGCTCTGTTTGTAT 276

Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60

QY 277 GCCATTCGGCGCTGATCCCACTGAAGCAGCAGCTGGAATATGATCAGCTGACCCCGCGG 336

Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80

QY 337 CGCTCCAGGAAGCTGAAGAGGTCGCTGTGACCGCTGTGACCCCGGAGCGCTCGGCGCTG 396

Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100

QY 397 AGTGTGCGTGGTGGCGCTGGAGTTGGCTGTGGCTCTTCATCTCCCACTCATCAAGGC 456

Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120

QY 457 GGTCCAGCAGACAGCTCGGCTCCAGGTAGGGGACGAGATCGTCCGGATCAATGGATAT 516

Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspIleValArgIleAsnGlyTyr 140

QY 517 TCCATCTCCTCTGTACCCCATGAGGAGGTGCATCAACCTCATTTGCAACCAAGAAACTGTG 576

Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160

QY 577 TCCATCAAGTGAGACACATCGGCGCTGATCCCGTGAAAGCTCTCTCTGATGAGCCCTC 636

Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180

QY 637 ACTTGGCAGTATGTGGATCAGTTTGTCTCGGAATCTGGGGCGCTGCGAGCAGCCCTGGC 696

Db 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200

QY 697 TCCTCTGGAAATCGGAAACAGAGAGAGAGGTCTTCATCAGCCTGGTAGGCTCCCGA 756
DB 201 SerProGlyAsnArgGluAsnLysGluLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCCTTGGCTGACGATTTCCAGCGCCGCCCATCCAGAGCCTGGCATCTTTATCAGCCAT 816
DB 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTGAACCTGGCTCCTCTGCTGCTGAGTGGGATGGAGATAGGAGCAGATTCGAA 876
DB 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCATGGCTGCTCTTCTTAACCTGGATCACAGGAGGCTGTAATGCTGCTGAAAT 936
DB 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
QY 937 AGCCGAGCCTGACCATCTCCATTGTAGCTGCAGCTGGCCGGAGCTGTTCATGACAG 996
DB 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGGAGCGCTGGCAGAGCGCGGAGCGTGAGCTGCAGCGCAGAGCTTCTCATGCGAG 1056
DB 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
QY 1057 AAGCGCTGGCGATGGAGTCCCAACAGATCCCTCCAGGAGCAGCAGAGATGGAGCGCAA 1116
DB 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln 340
QY 1117 AGGAGAAAGAAATGCCAGAGCAGCAGCAGAGGAGAAATGAGAGATACCGGAGGAGATG 1176
DB 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
QY 1177 GRACAGATTGTAGAGGAGGAGAGAGATTGTAGACCAATGGAGCAAGCACTGGGCTCA 1236
DB 361 GlnGlnIleValGlnGluGluLysPheLysLysGlnTrpGluAspTrpGlySer 380
QY 1237 AAGGAACAGCTACTCTGTGCTTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGC 1296
DB 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCCA----- 1302
DB 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
QY 1303 -----AAGTGATTTCCGGAATATGAGAAAGGCTTTGACCCCTA 1341
DB 421 GlyThrGluGluGlnGlyGluGln-AspPheArgLysTyrGluGluGlyPheAspProty 440
QY 1342 CTCTATGTTACCCAGAGCAGATCATGCGGAAGGATGTCGGCTCCTACGCATCAAGAA 1401
DB 440 rSerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLys 460
QY 1402 GGAGGGATCTTAGACCTGGCCCTGGAGCGGTGGACTCCCGCATTTGGGAGGTGT 1461
DB 460 sGluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProIleGlyLysValVa 480
QY 1462 CGTTTCTGCTGTATGAGCGGGAGCTGCTGAGCGGCATGTTGGCATTTGTAAGGGGA 1521
DB 480 lValSerAlaValTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAs 500
QY 1522 CGAGATCATGGCAATCAACGGCAAGATTGTACAGACTACACCCCTGGCTGAGCTGACGC 1581
DB 500 pGluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAl 520
QY 1582 TGGCCTCAGAGCCTGGAAATCAGCGCGGGAGTGGATCGACCTGTGTGTTGCCGTCTG 1641
DB 520 aAlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValValAlaValCy 540
QY 1642 CCCCCAAAGGATATCAGCATGAGCTGCTTCTTCTCAAGTCCCAAGGGAACCA 1701
DB 540 sProProLysGluTyrAspAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnG 560

QY 1702 AATTCAGCGCTTAGGAACACAGTGTAGCTCGGCCCCACCTCGTGAACACAAAGCCTCGAC 1761
DB 560 nIleHisAlaLeuGlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgTh 580
QY 1762 CAGCCTTGAGAGAGCCACATGACACACACAGATGCGATCCTTGGACCTGAATCTATC 1821
DB 580 rSerLeuGluArgGlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSe 600
QY 1822 ACCAGGAATCTCAACTCCCTTTGGCCTGAACCCAGCGCCAGATAAGGAACAGCTCGGG 1881
DB 600 rProArgAsnLeuLysLeuProLeuAlaLeuAsnGlnGlnIleArgAsnSerSerG 620
QY 1882 CCACCTTTTGAAGGCCAATCTGGAGGAAGGAGCAGCCGCTTTGGGAGAGATCT 1941
DB 620 yHisPhePheGluGlyGlnCysGlyGlyLysGlyAlaAlaSerArgLeuGlyGluAspLe 640
QY 1942 CAAGGATCCACACTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1978
DB 640 uLysAspProAspSerHisSerPheProLeuAlaGln 652
RESULT 3
AAY07098
ID AAY07098 standard; Protein; 481 AA.
XX AC AAY07098;
XX DT 02-JUL-1999 (first entry)
XX DE Colon cancer associated antigen precursor sequence.
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
XX KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX KW prostate cancer.
XX OS Homo sapiens.
XX PN WO9904265-A2.
XX PD 28-JAN-1999.
XX PF 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
XX PR 10-OCT-1997; 97US-0061599.
XX PR 10-OCT-1997; 97US-0061765.
XX PR 10-OCT-1997; 97US-0948705.
XX PR 11-OCT-1997; 97GB-0021697.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX DR WPI; 1999-132448/11.
XX PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX PS Disclosure: Page 664-665; 787pp; English.
XX CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.

CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

Sequence	481 AA;
SQ	

Alignment Scores:

Pred. No.:	7,41e-181	Length:	481
Score:	2188.00	Matches:	438
Percent Similarity:	91.06%	Conservative:	0
Best Local Similarity:	91.06%	Mismatches:	1
Query Match:	54.85%	Indels:	43
DB:	20	Gaps:	1

US-09-502-945-5 (1-2162) x AAY07098 (1-481)

QY	97	ATGGACCGAAAGTGGCCCGAGAAATCCGGGCATCAAGGTGGATTTCTGATTTGAAATGAT	156
Db	1	MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp	20
QY	157	GCAGAGAAGGACATCTCTATGATGTGCTCGGAATGTACCACAGACCATGGAGCGGCC	216
Db	21	AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla	40
QY	217	GTGCTCTGGGAGACCTGAAGCTGGTCATCAATGAACCCAGCCGCTGTGCTCTGTTTGAT	276
Db	41	ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp	60
QY	277	GCCATTGGCCCGCTGTATCCACATGAAGCACCAGGTGGATATCATCAGTGCACCCCGG	336
Db	61	AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg	80
QY	337	CGCTCCAGGAAGCTGAAGGAGGTGCGCTCTGGACCGTCTGCACCCCGAAGGCTCTGGCCGTG	396
Db	81	ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu	100
QY	397	AGTGTGCGTGTGGCTGTGGAGTTTGCTGTGGGCTCTTCATCTCCACCTCATCAAGGC	456
Db	101	SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly	120
QY	457	GGTCAGGCAGACAGCTCGGGCTCCAGGTAGGGAGGAGAGATCCTCGGGATCAATGATAT	516
Db	121	GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr	140
QY	517	TCCATCTCCTCTGATACCATGAGGAGGTGCATCAAGCTCATTCGAAACCAAGAAACTGTG	576
Db	141	SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal	160
QY	577	TCCATCAAGTGAGACACATCGGCCCTGATCCCGCTGAAAGCTCTCTCTGATGAGCCCTTC	636
Db	161	SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu	180
QY	637	ACTTGGCAGTATGTGATCAGTTTGTCTGGAAATCTGGGGCTGTGGAGCGAGCGACCTGGCG	696
Db	181	ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlySerLeuGly	200
QY	697	TCCCTTGGAAATCGGGAACAAGGAGAGAAGGTCTTCATCAGCCTGTGAGCTCCCGA	756
Db	201	SerProGlyAsnArgGluAsnLysGluLysLysValPheIleSerLeuValGlySerArg	220
QY	757	GGCCTTGGCTGCAGCATTTCCAGCGGCCCATCCACAAGCCGTGGCATCTTTATCAGCCAT	816
Db	221	GlyLeuGlyCysSerIleSerGlyProIleGlnLysProGlyIlePheIleSerHis	240
QY	817	GTGAACCTGGCTCCCTGTCTGTGAGGTGGGATTTGGAGATAGGGGACACAGATTGTGAA	876
Db	241	ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu	260
QY	877	GTCAATGGGCTGCACCTTCTTACCTGGATCACAGAGGCTGTAAATGTGCTGAAATAT	936
Db	261	ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn	280

QY	937	AGCCGACGCTGACCATCTCCATTGTAGCTGACGTGCGGGAGAGCTGTTCATGACAGAC	999
Db	281	SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp	300
QY	997	CGGAGCGGCTGGCAGAGGCGCGCAGCTGACCTCCAGCGCAGAGCTTCTCATGCAG	1056
Db	301	ArgGluArgLeuAlaGluAlaAa-rGgInArgGluLeuGlnArgGlnLeuLeuMetGln	320
QY	1057	AAGCGGCTGGCGATGGAGTCCAAACAAGATCCTCCAGGACGACGAGAGATGGAGCGGCAA	1111
Db	321	LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnGlnGlnGlnGln	340
QY	1117	AGGACAAAGAAATGCCCCAGAGGACGACAGAGAAATACAGAGATACCCGGAAGGAGATG	1177
Db	341	ArgArgLysGluIleAlaGlnLysAlaAlaGluAsnGluArgTyrArgLysGluMet	360
QY	1177	GAACAGATTGAGAGGAGGAGAGAGTTTAAGAAGCAATGGGAAGAAGACTGGGGCTCA	1233
Db	361	GluGlnIleValGluGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer	380
QY	1237	AAGGAACAGCTACTCTTGGCCATAAACCATCACCTGCTGAGGTACACCCAGTACCCCTTCGC	1299
Db	381	LysGluGlnLeuLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg	400
QY	1297	AAGCCAAAGTG-----	1307
Db	401	LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly	420
QY	1307	-----	1307
Db	421	GlyThrGluGluGlnGlyGluGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	440
QY	1308	-----ATTTCGGAAATATAGGAAGGCTTTGACCCCTACTCTATGT	1344
Db	441	AspSerIleGlnAspLysIleSerGlyAsnMetArgLysAlaLeuThrProThrLeuCys	460
QY	1350	TCACCCGACGACATCATGGGAAGGATGTCCGGCTCCTAGCGATCAAGAAGGAGGAT	1405
Db	461	SerProGlnSerArgSerTrpGlyArgMetSerGlySerTyrAlaSerArgArgArgAsp	480
QY	1410	CCT 1412	
Db	481	Pro 481	
RESULT	4		
AY07099			
ID	AY07099	standard; Protein; 521 AA.	
XX	AC	AY07099;	
XX	DT	02-JUL-1999 (first entry)	
XX	DE	Colon cancer associated antigen precursor sequence.	
XX	KW	Cancer associated antigen; diagnosis; research; treatment; human;	
XX	KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
XX	KW	prostate cancer.	
OS		Homo sapiens.	
XX	FN	W09904265-A2.	
XX	PD	28-JAN-1999.	
XX	PF	15-JUL-1998; 98W0-US14679.	
XX	PR	22-JUN-1998; 98US-0102322.	
PR	PR	17-JUL-1997; 97US-0896164.	
PR	PR	10-OCT-1997; 97US-0061599.	
PR	PR	10-OCT-1997; 97US-0061765.	
PR	PR	10-OCT-1997; 97US-0948705.	
PR	PR	11-OCT-1997; 97GB-0021697.	

PA (LUDW-) LUDWIG INST CANCER RES.
 XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
 PI Pfeunderschuh M, Sahin U, Scanlan MJ, Stockert E;
 PI Tureci O;
 XX WPI; 1999-132448/11.
 XX New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 XX
 PS Disclosure: Page 666-667; 787pp; English.
 XX
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 XX
 SQ Sequence 521 AA;
 Alignment Scores:
 Pred. No.: 4, 18e-179 Length: 521
 Score: 2168.00 Matches: 438
 Percent Similarity: 84.07% Conservative: 0
 Best Local Similarity: 84.07% Mismatches: 1
 Query Match: 54.35% Indels: 83
 DB: 20 Gaps: 1
 US-09-502-945-5 (1-2162) x AAY07099 (1-521)
 QY 97 ATGACCGAAAGTGGCCGAGAAATCCGCGATAGTGGATTCTCTGATGAAATGAT 156
 Db 1 MCAAPARGLYSVALAARGGLUPHEARGHISLYSVALASPHELEULEGLUASNASP 20
 QY 157 GCAGAGAAGGACTATCTCTATGATGTGCTGCGGAATGTACCACGACACCATGGACGTGGCC 216
 Db 21 ALAAGLULYSASPTYRLEUTYRASPVALLEUARGMETYRHSGLNTHRMETASPVALALA 40
 QY 217 GTGCTGTGGGACCTGACCTGAAGCTGGTCATCAATGAACCCAGCCGCTGCTCTGTTGAT 276
 Db 41 VALLEUVALGLYASPLEULYSLEUVALILEASNGLUPROSERARGLEUPROLEUPHEASP 60
 QY 277 GCATTCGGCGGTGATCCCACTGAAGCACCAGGTGGATATGATCAGTGCACCCCGG 336
 Db 61 ALAILEARGPROLEULEPROLEULYSGLNVALGLUTYRASPGLNLEUTHRPROARG 80
 QY 337 CGCTCCAGGAAGCTGAAGAGGTGCGTCTGACCGCTGTGACCCCGAAGCCCTCGCGCTG 396
 Db 81 ARGSERARGLYSLEULYSGLUVALARGLEUASPARGLEUHSISPROGLUGLYLEUGLYLEU 100
 QY 397 AGTGTGCGTGGTGGCTGGATTGGCTGTGGCTCTCTCATCTCCACCTCATCAAGGC 456
 Db 101 SERVALARGGLYGLYLEUGLUPHEGLYCYSGLYLEUPHEILESERHISLEULELYSGLY 120
 QY 457 GGTGACGACAGCGTGGGCTCCAGGTAGGGAGCAGATCGCTCGGATCAATGATAT 516
 Db 121 GLYGLNALAASPVALGLYLEUGLNVALLGLYASPLUULEVALARGILEASNGLYTYR 140
 QY 517 TCCATCTCTCTCTACCCATGAGGAGTTCATCAACTCATTCGACCCAGAAACCTGG 576
 Db 141 SERILESERCYSTRHISGLUGLUVALILEASNLEULEARGTHRILYSLYSTHRVAL 160

QY 577 TCCATCAAGTGAACACATCGCGCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTC 636
 Db 161 SERILELYSVALARGHISILEGLYLEUILEPROVALYSSERSERPROASPGLEUPROLEU 180
 QY 637 ACTTGGCAGTATGTGATCAGTTTGTGCGAATCTGGGGCGTGGCAGGCGCTGGGC 696
 Db 181 THRTRPGLNTRYVALASPGLNPHEVALSERGLUSERGLYGLYVALARGGLYSERLEUGLY 200
 QY 697 TCCCTCGAAATCGGAAACAAAGGAGAAGTCTTCATCAGCCCTGGTAGGCTCCCGA 756
 Db 201 SERPROGLYASNARGGLUASNLYSLYSVALPHEILESERLEUVALGLYSERARG 220
 QY 757 GGCCTTGGCTGCAGCATTTCCAGCGGCCCATCCAGAGCCTGGCATCTTATCAGCCAT 816
 Db 221 GLYLEUGLYCYSERILESERSERGLYPROFLEGLNLYSPROGLYILEPHEILESERHIS 240
 QY 817 GTGAAACCTGCTCCCTGCTCTGCTGAGTGGATTGAGATGAGGACCGACCATGTCTGAA 876
 Db 241 VALLYSPROGLYSERLEUSERALAGLUVALGLYLEUGLULEGLYASPLNLEUVALGLU 260
 QY 877 GTCAATGCGCTGACTTCTTAACTGGATCACAAGGAGGCTGTAATGTGCTGAAAT 936
 Db 261 VALASNGLYVALASPHESERASNLEUASNLYSLYSGLUVALASNVALLEULYSASN 280
 QY 937 AGCCGACCTGACCATCTCCATTGTAGCTGCAGCTGCCCGGAGCTGTTTCATGACAGAC 996
 Db 281 SERARGSERLEUTHRILESERILEVALALAALAAGLYARGGLULEUPHEMETHRASP 300
 QY 997 CGGAGCGCTGGCAGAGCGCGCTGAGCTGCAGCGCAGGAGGAGCTTCTCATGCGAG 1056
 Db 301 ARGGLUARGLEUALAAGLUALAARGLNARGGLULEUGLNARGGLULEULEUMETGLN 320
 QY 1057 AAGCGCTGGCGATGGAGTCCCAAGATCCTCCAGGACGACGAGAGATGAGGCGGCAA 1116
 Db 321 LYSARGLEUALAMETGLUSERASNLYSILELEUGLNGLUGLNGLUMETGLUARGLN 340
 QY 1117 AGGAGAAAGAAATGCCCCAGAGGACGACGAGGAAATGAGATATCCGGAAGGAGATG 1176
 Db 341 ARGARGLYSGLUILEALAGLNLYSALAAGLUASNGLUARGTYRARGLYSGLU 360
 QY 1177 GAACAGATTGTAGAGGAGGAAGAGAAAGTTTAAAGCAATGGGAAGAGCTGGGCTCA 1236
 Db 361 GLUGLNILEVALGLUGLUGLULYSLEULYSGLINTRPGLUUASPTRPGLYSER 380
 QY 1237 AAGGAACAGCTACTCTTCCCTAAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGC 1296
 Db 381 LYSGLUGLNLEULEUPROLYSTHRIETHRALAGLUVALHISPROVALPROLEUARG 400
 QY 1297 AAGCCAAAGTG----- 1307
 Db 401 LYSPROLYSTYRASPGLNGLYVALGLUPROGLULEUGLUPROALAASPSPLEUASPGLY 420
 QY 1307 ----- 1307
 Db 421 GLYTHRGLUGLUGLNGLYGLUINLTHRPHYCYSPROSERPROGLNPROPROARGGLYPRO 440
 QY 1307 ----- 1307
 Db 441 GLYVALSERTHRILESERLYSPROVALMETVALHISGLINGLUPROASNPHILETYRARG 460
 QY 1307 ----- 1307
 Db 461 PROALAVALLYSSERGLUVALLEUPROGLNGLUMETLEULYSARGMETVALVALTYRGLN 480
 QY 1308 -----ATTTCCGGAATATGAGAAAGCTTTTCACCCCTACTCTATGT 1349
 Db 481 ASPSERILEGLINASPGLYSILESERGLYASNMETARGLYSALALETURHLEUTHRLEUCYS 500
 QY 1350 TCACCCCGACAGCAGATCATGCGGAAGGATGTCGGCTCCTACGCAATCAAGAGGAGGAT 1409
 Db 501 SERPROGLNINSEARGSERTRPGLYARGMETSERGLYSERTYRALASERARGARGASP 520
 QY 1410 CCT 1412


```

Db 521 Pro 521
RESULT 5
ID AAY07095 standard; Protein; 403 AA.
XX AC AAY07095;
XX DT 02-JUL-1999 (first entry)
XX DE Colon cancer associated antigen precursor sequence.
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX OS Homo sapiens.
XX PN W09904265-A2.
XX PD 28-JAN-1999.
XX PF 15-JUL-1998; 98WO-US14679.
XX PR 22-JUN-1998; 98US-0102322.
XX PR 17-JUL-1997; 97US-0896164.
XX PR 10-OCT-1997; 97US-0061599.
XX PR 10-OCT-1997; 97US-0061765.
XX PR 10-OCT-1997; 97US-0948705.
XX PR 11-OCT-1997; 97GB-0021697.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX DR WPT; 1999-132448/11.
XX PT New isolated cancer associated nucleic acids and polypeptides -
XX PT isolated using sera from cancer patients, used to develop products
XX PT for the diagnosis, monitoring or treatment of cancers
XX PS Disclosure; Page 659-660; 787pp; English.
XX CC The invention relates to a method for diagnosing a disorder characterised
XX CC by expression of a human cancer associated antigen precursor coded for by
XX CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX CC biological sample isolated from a subject with an agent that specifically
XX CC binds to the NAM, an expression product or a fragment of an expression
XX CC product complexed with an HLA molecule; and (b) determining the
XX CC interaction between the agent and the NAM or the expression product as a
XX CC determination of the disorder. The products and methods can be used in
XX CC the diagnosis, monitoring, research, or treatment of conditions
XX CC characterised by the expression of various cancer associated antigens.
XX CC The invention provides nucleic acid sequences and encoded polypeptides
XX CC which are cancer associated antigen precursors expressed in human breast
XX CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX CC lung cancer.
XX SQ Sequence 403 AA;

Alignment Scores:
Pred. No.: 1.42e-168 Length: 403
Score: 2046.00 Matches: 403
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 51.29% Indels: 0
DB: 20 Gaps: 0

US-09-502-945-5 (1-2162) x AAY07095 (1-403)

```

```

QY 97 ATGGACCGAAAAGTGGCCCGAGGATCCCGCATAGAGTGGATTCTCTGATTGAAATGAT 156
Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
QY 157 GCAGAGAAGGACTATCTCTATGCTGCTGCAATGTACCACAGACCATGGACGTGGCC 216
Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
QY 217 GTGCTCGTGGAGACCTGAAGCTGGTCATCAATGAACCCAGCCGCTCTCTCTGTTGAT 276
Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
QY 277 GCATTCCGGCGCTGATCCCACTCAGACCAAGTGGCAATATCATCAGCTGACCCCGG 336
Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
QY 337 CGCTCCAGGAAGCTGAAGAGGTCCGCTCTGGACCGCTGCACCCCGAAGCCCTCGGCTG 396
Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
QY 397 AGTGTGCTGGTGGCTGGAGTTTGGCTGTGGGCTCTTCATCTCCCACTCATCAAGGC 456
Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
QY 457 GGTCAAGCAGACACCGCTCGGCTCCAGGTAGGGAGGACGAGATCGTCCGGATCAATGAT 516
Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
QY 517 TCCATCTCCTCTCTACCCATGAGAGGTTCATCAACTCATTCGAAACCAAGAAACTGTG 576
Db 141 SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
QY 577 TCCATCAAGTGAGACACATCGCCCTGATCCCGTGAAGAGCTCTCTGTATGAGCCCTC 636
Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProAspGluProLeu 180
QY 637 ACTTGGCAGTATGTGATCAGTTTGTGCGAATCTGGGGCTGGGGAGGACCCCTGGGC 696
Db 181 ThrTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlySerLeuGly 200
QY 697 TCCCTCGAAATCGGAAAAAAGAGAGAGTCTTCATCAGCTGTGTAGGCTCCCGA 756
Db 201 SerProGlyAsnArgGluAsnLysGluLysValPheIleSerLeuValGlySerArg 220
QY 757 GGCTTGGCTGACGATTCACGGCCCATCCAGAGCCCTGGCATCTTTATCAGCCAT 816
Db 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
QY 817 GTAAACCTCGCTCCCTCTCTGCTGAGTGGGATTTGAGATAGGGACGAGATTGTCGAA 876
Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
QY 877 GTCAATGGCTCGCACTTCTTAACCTGGATCACAAGAGGCTGTAATGTGCTGAAAAAT 936
Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
QY 937 AGCGCAGCTGACCATCTCCATTTAGCTGACGTGGCGGAGCTGTTTCATGACAGAC 996
Db 281 SerArgSerLeuThrIleSerIleValAlaAlaGlyArgGluLeuPheMetThrAsp 300
QY 997 CGGGAGCGGCTGGCAGAGCGCGGAGCTGACGTGCAGCGGAGGAGCTTCTCATGCGAG 1056
Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnLeuLeuMetGln 320
QY 1057 AAGCGCTGCGGATGGAGTCCAAAGATCCTCCAGGAGCAGGAGGATGGAGCGGCAA 1116
Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGlnGlnGlnMetGluArgGln 340
QY 1117 AGGAGAAAAGAAATTCGCCAAGAGCGCAGAGAGAAATGAGAGATACCGGAGGAGATG 1176
Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
QY 1177 GACAGATTCTAGAGGAGGAGAGAGATTAAAGCAATGGGAAGAGACTGGGCTCA 1236

```



```
Db 361 GluGlnIleValGluGluGluLysPheLysGlnTrpGluGluAspTrpGlySer 380
QY 1237 AAGGAACAGCTACTTCCCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGC 1296
Db 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
QY 1297 AAGCCAAAG 1305
Db 401 LysProLys 403
RESULT 6
RAY07097
ID AAY07097 standard; Protein: 328 AA.
AC AAY07097;
XX 02-JUL-1999 (first entry)
XX Colon cancer associated antigen precursor sequence.
DE Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX Homo sapiens.
OS
PN W09904265-A2.
XX 28-JAN-1999.
XX 15-JUL-1998; 98WO-US14679.
XX 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Disclosure; Page 662-663; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 328 AA;
Alignment Scores:
Pred. No.: 3.09e-126 Length: 328
```

```
Score: 1557.50 Matches: 302
Percent Similarity: 92.10% Conservative: 1
Best Local Similarity: 91.79% Mismatches: 0
Query Match: 39.04% Indels: 26
DB: 20 Gaps: 1
US-09-502-945-5 (1-2162) x AAY07097 (1-328)
QY 1069 ATGGAGTCCAAACAGATCCTCCAGGAGCAGAGATGAGCGGCAAGAGAGAAAGAA 1128
Db 1 MetGluSerAsnLysIleLeuGlnGluGlnMetGluArgGlnArgArgLysGlu 20
QY 1129 ATTCCCCAGAGGCAGCAGAGGAAATGAGATACCGGAGGAGATGGACACATTGTA 1188
Db 21 IleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMetGluGlnIleVal 40
QY 1189 GAGGAGGAAGAGAAGTTTAAAGAACAATGGGAAGAAGACTGGGGCTCAAGGAACAGCTA 1248
Db 41 GluGluGluLysPheLysLysGlnTrpGluLysTrpGlySerLysGluGlnLeu 60
QY 1249 CTCTTGCTTAAACCATCATCTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCA----- 1302
Db 61 LeuLeuProLysThrIleThrAlaGluValHisProValProLeuArgLysProLysTyr 80
QY 1302 ----- 1302
Db 81 AspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGlyGlyThrGluGlu 100
QY 1303 -----AAGTGATTTCCGGAATATGAGGAAGGCTTTGACCCCTACTCTATTGTTCAAC 1353
Db 101 GlnGlyGluGln-AspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheTh 120
QY 1354 CCCAGAGCAGATCATGGGAAGGATGTCGGCTCTCTAGCGATCAAGAGGAGGATCCTT 1413
Db 120 rProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLe 140
QY 1414 AGACTGCGCTTGGGAAGCGGTGGTGGCTCCCATCGGAAGGTGGTCTGCTGTGT 1473
Db 140 uAspLeuAlaLeuGluGlyValAspSerProIleGlyLysValValSerAlaVal 160
QY 1474 GTATGAGCGGGAGCTGCTGAGCGCATGTGTGGCATTTGTAAAGGGGACGAGATCATGGC 1533
Db 160 lTyrGluArgGlyAlaAlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAl 180
QY 1534 AATCAACGGCAAGATGTGACAGACTACACCTGGCTGAGGCTGACGCTGCCCTGCAGAA 1593
Db 180 alleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAlaLaLeuGlnLy 200
QY 1594 GGCCTGGATCAGGGCGGGAGTGGATCGACTGTGTGGTTCGCTGCCCTCCCAAGGA 1653
Db 200 sAlaTrpAsnGlnGlyGlyAspTrpIleAspLeuValValAlaValCysProProLysGl 220
QY 1654 GTATGACGATGAGCTGACCTTCTTGTGAGTCCAAAGGGGAAACCAATTCACGCGTT 1713
Db 220 uTyrAspAspGluLeuThrPheLeuLeuLysSerLysArgGlyAsnGlnIleHisAlaLe 240
QY 1714 AGGAACAGTGTGAGTCCGGCCCTCCCTCGTGAACACAAAGCTCGGACCAAGCTTGAGAG 1773
Db 240 uGlyAsnSerGluLeuArgProHisLeuValAsnThrLysProArgThrSerLeuGluArg 260
QY 1774 AGGCCATGTACACACACACATGCGATCCTTGGGACCTGAATCTATCACCACGGAATCT 1833
Db 260 gGlyHisMetThrHisThrArgTrpHisProTrpAspLeuAsnLeuSerProArgAsnLe 280
QY 1834 CAACCTCCCTTTGGCCCTGAACACCGGCGCAGATGAAGAACAGCTCGGCGCCACTTTTGA 1893
Db 280 uLysLeuProLeuAlaLeuAsnGlnGlyIleArgAsnSerSerGlyHisPhePheGl 300
QY 1894 AGGCCAATGTGGAGAAAGGAGGAGCAGCCGCTTTGGGAGAGATCTCAGGATCCAGA 1953
Db 300 uGlyGlnCysGlyGlyLysGlyAlaAlaSerArgLeuGlyGluAspLeuLysAspProAs 320
QY 1954 CTCATTCCTTCTCTGGCCCGAG 1978
```


Db 320 pSerHisSerPheProLeuAlaGln 328
 RESULT 7
 AAY38440
 ID AAY38440 standard; Protein; 268 AA.
 XX
 AC AAY38440;
 XX
 DT 18-OCT-1999 (first entry)
 XX
 DE Human secreted protein encoded by gene No. 5.
 XX
 KW Human: secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 PN WO9935158-A1.
 XX
 PD 15-JUL-1999.
 XX
 PF 06-JAN-1999; 99WO-US00108.
 XX
 PR 07-JAN-1998; 98US-0070704.
 PR 07-JAN-1998; 98US-0070657.
 PR 07-JAN-1998; 98US-0070658.
 PR 07-JAN-1998; 98US-0070692.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Duan RD, Ebner R, Lafleur DW, NI J;
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
 XX
 DR WPI: 1999-444190/37.
 DR N-PSDB; AA206223.
 XX
 PT New isolated human genes and the secreted polypeptides they encode
 PS
 PS Disclosure: Page 206-207; 227pp; English.
 CC
 CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AA206210) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 36 novel genes and their fragments (nucleic
 CC acid sequences: AA206219-206263; amino acid sequences AAY38386-Y38498)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 36
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AA206219 for described uses).
 XX
 SQ Sequence 268 AA;

Alignment Scores:
 Pred. No.: 3,87e-93 Length: 268
 Score: 1175.50 Matches: 236
 Percent Similarity: 90.11% Conservatives: 1
 Best Local Similarity: 89.73% Mismatches: 0
 Query Match: 29.47% Indels: 26
 DB: 20 Gaps: 1

US-09-502-945-5 (1-2162) x AAY38440 (1-268)

QY 964 GCTGACAGCTGGCGGAGCTGTTTCATGACAGACCGGAGCGGCTGCAGAGCGCGCAG 1023
 Db 6 AlaAlaAlaGlyArgGluLeuPheMetThrAspArgGluArgLeuAlaGluAlaArgGln 25
 QY 1024 CGTGAGCTGCAGCGGAGGAGCTTCTCATGCGAGAACCGGCTGGCGATGGAGTCCACACAG 1083
 Db 26 ArgGluLeuGlnArgGlnGluLeuMetGlnLysArgLeuAlaMetGluSerAsnLys 45
 QY 1084 ATCCTCCAGGAGCAGCAGGAGATGGAGCGGCAAGAGAGAAAATAATGCCAGAGGCA 1143
 Db 46 IleLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 65
 QY 1144 GCAGAGAAATAGAGATACCGGAGAGATGGAACAGATGTAGAGAGAGAGAGAG 1203
 Db 66 AlaGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 85
 QY 1204 TTTAACAACAATGGGAACAAGCTGGGCTCAAGAGACACACTACTCTTGCCTAAACAC 1263
 Db 86 PheLysLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 105
 QY 1264 ATCCTGCTGAGTACACCCAGTACCCCTTCGCAAGCCA----- 1302
 Db 106 IleThrAlaGluValHisProValProLeuArgLysProLysTyraAspGlnGlyValGlu 125
 QY 1303 -----AAGTGA 1308
 Db 126 ProGluLeuGluProAlaAspAspLeuAspGlyGlyThrGluGluGlnGlyGluGln-As 145
 QY 1309 TTTCCGGAATATGAGGAGGCTTTGACCCCTACTCTATGTTCCACCCAGACAGATCAT 1368
 Db 145 pPheArgLysTyraGluGluGlyPheAspProtyrSerMetPheThrProGluGlnIleMe 165
 QY 1369 GGGGAAGGATGTCGGCTCTACGATCAAGAGAGGAGGATCTTAGACCTGGCCCTGGA 1428
 Db 165 TGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeuAspLeuAlaLeuGln 185
 QY 1429 AGCGGTGTGGACTCCCGCTCCCGCTGGAGGTGTGCTGTTCTGCTGTATGAGCGGGAGC 1488
 Db 185 uGlyGlyValAspSerProIleGlyLysValValSerAlaValTyraGlyAla 205
 QY 1489 TGCTGAGCGGATGTTGGCTTGTGAAGGGGAGGAGATCATGGCAATCAACGGCAAGAT 1548
 Db 205 alaGluArgHisGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLysIle 225
 QY 1549 TGTGACAGACTACACCTCGCTGAGCTGAGCTGCGCTGCGCTGCAGAGAGCGCTGGAATCAGG 1608
 Db 225 eValThrAspTyraThrLeuAlaGluAlaAspAlaAlaLeuGlnLysAlaTrpAsnGlnGln 245
 QY 1609 CGGGGACTGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
 Db 245 yGlyAspTrpIleAspLeuValAlaValCysProLysGluTyraAspGluLe 265
 QY 1669 GACCTTC 1675
 Db 265 thrPhe 267
 RESULT 8
 AAB53238
 ID AAB53238 standard; Protein; 120 AA.
 XX
 AC AAB53238;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human colon cancer antigen protein sequence SEQ ID NO:778.
 XX
 KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antineoplastic; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;

KW infectious disease; cardiovascular disorder.

OS Homo sapiens.

XX W020005351-AL.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA,* Ruben SM;

XX WPI; 2000-587534/55.

DR N-PSDB; AAC97995.

XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer

XX Claim 11; Page 1339; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and
CC vulnary, nephrotropic, antineoplastic and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.

XX SQ Sequence 120 AA;

Alignment Scores:
Pred. No.: 4.62e-29 Length: 120
Score: 435.00 Matches: 97
Percent Similarity: 80.65% Conservative: 3
Best Local Similarity: 78.23% Mismatches: 14
Query Match: 10.90% Indels: 10
DB: 21 Gaps: 4

US-09-502-945-5 (1-2162) x AAB53238 (1-120)

QY 1196 AAGAGAGTTTAAGA---AGCAATCGGAGAGAGATGGGCTCAAGAGACAGCTACTCT 1252

Db 6 GluArgHisLeuArgValThrGlnGlyLeuGlyThrGlyAlaPheLeuGlyGly----- 23

QY 1253 TGCCTAAACACTACTCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTGATTC 1312

Db 24 ---LeuArgProValLeu-----GlnProArgGlnGlyGln---AspPhe 36

QY 1313 CGGAATATGAGAGGCTTTGACCCCTACTCTATGTTACACCCAGAGCAGATCATGGGG 1372

Db 37 ArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGluGlnIleMetGly 56

QY 1373 AAGATGTCGGCTCCTACGCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAAGGC 1432

Db 57 LysAspValArgLeuArgIleLysLysGluGlySerLeuAspLeuAlaLeuGlyGly 76

QY 1433 GTGTGGACTCCCATTTGGGAGGTGGTCTGTTCTGCTGTGTATGACGGGGAGCTGCT 1492

Db 77 GlyValAspSer***IleGlyLysValValSerAlaValTyrGluArgGlyAlaAla 96

QY 1493 GAGCGCATGGTGCATTTGTGAAGGGGACGAGATCATGCAATCAACGGCAAGATTGTG 1552
Db 97 GluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLysIleVal 116

QY 1553 ACAGACTACACC 1564

Db 117 ThrAspTyrThr 120

RESULT 9

ABB61924

ID ABB61924 standard; Protein; 493 AA.

XX AC

XX ABB61924;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 12564.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL06027.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 12564; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS7072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 493 AA;

Alignment Scores:
Pred. No.: 2.66e-27 Length: 493
Score: 418.00 Matches: 128
Percent Similarity: 45.89% Conservative: 62
Best Local Similarity: 30.92% Mismatches: 157
Query Match: 10.48% Indels: 67
DB: 22 Gaps: 12

US-09-502-945-5 (1-2162) x ABB61924 (1-493)

QY 340 TCCAGGAGCTGAGGAGGTGCGCTCTGGACGCTGCACCCCGAAGGCCTC----- 390

Db 58 SerSerArgLeuArgValLeuArgLeuValArgProHisGlnArgArgLeuValGly 77

QY	391	-----GGCTCAGTGTGCGTGTCGGCTGCAGATTTGGCTGT	426
Db	78	GlyProGluArgGlySerThrTyrglyPheThrValargglycylargGluHisGlyThr	97
QY	427	GGGCTCTTCATCTCCACCTCATCAAGGGCGTTCAGGCAGACAGCGTGGCTCCAGGTA	486
Db	98	GlyPheValSerHisValGluHisGlyGlyGluAlaHisLeuLysSgLyLeuArgile	117
QY	487	GGGACGAGATCGCCGGATCAATGATATTCCATCTCTCTGTACCCTAGAGAGTC	546
Db	118	GlyaspGlnIleuArgIleAsnGlyPheArgLeuAspAlaValHisLysGluPhe	137
QY	547	ATCAACTCATTCGAACAAGAAAGTGTCTCATCAAAGTCAGACACATGGCTGATC	606
Db	138	IleGlnLeuValAlaGlyGlnaspargvalThrLeuLysValargGlyValGlyMetLeu	157
QY	607	CCCTGAAAAGCTCTCTGATGAGCCCCACTTGGCACGATATGTGCATCAGTTTGTGCG	666
Db	158	ProValArgaspLeuProGluGluArgLeuSerTrpSerValValLys-----LeuPro	175
QY	667	GAATCTGGGCGTGGCAGCCCTGGGCTCCCCTGGAAATCGGAAACACAGAGAAG	726
Db	176	SerValSerGlyThrProSerGluSerSerPheLysGlyGluArgArgGlyAlaSerArg	195
QY	727	AAGTCTTCATCAGCCCTG---GTAGGCTCCGA---GGCTTGGTGCAGCATTTCCAGC	780
Db	196	AspileSerValValLeuHisValAlaProArgThrLysLeuGlyLeuGlyIleCysLys	215
QY	781	GGCCCCATCAGAACCTGCCATCTTTATCAGCCATGTGAACCTCGCTCTGTCTGCT	840
Db	216	GlyProGluTrpLysProGlyIlePheValGlnPheThrLysAspArgSerValAlaArg	235
QY	841	GAGCTGGATGGAGATPAGGGACCATGTTCGAGTCAATGGCGTGCAGCTTCTCTAAC	900
Db	236	GluAlaGlyLeuargProGlyaspGlnIleLeuSerValAsnSerIleAspPheSerAsp	255
QY	901	CTGATCACAAGGAGGCTGTAATGTGCTGAAAAATAGCCGAGCGCTGACCATCTCCATT	960
Db	256	ValLeuPheSerGluAlaValAlaValMetLysSerSerLysLeuAspMetValVal	275
QY	961	GTAGCTCAGCTGCCGGAGCTGTTATGCATGACAGACCGGGAGCGCTGGCAGAGCGCGG	1020
Db	276	ArgThrAlaAlaGlyCysLeuPheProGlyGlu-SerSerGlyTyraSnSerSerAl	295
QY	1021	CAGCGTGCAGCTGCAGGAGGAGCTTCATGTCAGACGCGCTCGCATGGAGTCCAAC	1080
Db	295	aSer---SerValThrGlyaspGlnSerProcys-----TpAlaaspAl	309
QY	1081	AAGATCTCAGGACGACGAGGATGGAGCGGCAAGGAGAGAAAATAATGCCAGAAG	1140
Db	309	aLysSerLysArg-LeuThr-----AlavalargGluSerGlyalaglyGlyG	326
QY	1141	GCACGACGAAAATGAGAGATACCGAAGGAGATGGAAACAGATGTGTAGAGAGAGACAG	1200
Db	326	lyGlyCysGlyLeuSerSerAlaProGlyAlaGlySerProAsnTrpSerGlnGlyValG	346
QY	1201	AAGTTTAAGAAAGTATGGGAAGAGACTG-----	1229
Db	346	luValHisLysGlnMetAsnLysThreIleLysLeuThrGluAsnGlyThrSerIleA	366
QY	1230	-----GGGTCAAGGAA	1242
Db	366	sAsnThrTyrlleAlaSerThrGlyGlySerSerValSerGlySerGlySerThrGlyS	386
QY	1243	CAGTACTCTTGCTAAAACCATCACTGCTGAGGTACACCGAGTACCCCTWTCGCAAGCCA	1302
Db	386	erglyThrSerGlyArgSerGlnGlnSerGlnSerAsnProSerAsnProSerArgAsn-	405
QY	1303	AAGTGATTCCCGAAATATGAGAGGCT----TTGACCCCTACT-----	1343
Db	406	-----SerThrMetLysArgSerHisLeuArgProValAsnSerAlaGlySerG	423

QY	1344	-----CRATGTCACCCAGCAGCATCATGGGAAG-----ATGTCCGGCTCCT	1389
		::	
Db	423	lyIleGlyLeuSerSerGlySerAlaGlySerAlaGlySerAlaGlySerGlySerG	443
QY	1390	ACGCATCAAGAAGGAGGATCCTTACAGCCTGGCCCT	1425
		::: ::	
Db	443	lySer-ArgSerGlyGlyValIleAlaProAlaPro	454
RESULT 10			
AAY38390			
ID	AAY38390 standard; Protein; 98 AA.		
XX	AAY38390;		
DT	30-SEP-1999 (first entry)		
DE	Human secreted protein encoded by gene No. 5.		
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;		
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;		
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;		
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;		
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;		
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;		
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;		
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.		
OS	Homo sapiens.		
PN	WO9933158-A1.		
XX			
PD	15-JUL-1999.		
XX			
PF	06-JAN-1999; 99WO-US00108.		
XX			
PR	07-JAN-1998; 98US-0070704.		
PR	07-JAN-1998; 98US-0070657.		
PR	07-JAN-1998; 98US-0070658.		
PR	07-JAN-1998; 98US-0070692.		
XX			
FA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Brewer LA, Duan RD, Ebner R, Lafleur DW, NI J;		
PI	Olsen HS, Rosen CA, Ruben SM, Soppet DR;		
XX			
DR	WPI; 1999-444190/37.		
DR	N-PSDB; AAZ06223.		
PT	New isolated human genes and the secreted polypeptides they encode		
XX			
FS	Claim 11; Page 178-179; 227pp; English.		
XX			
CC	This sequence represents a secreted human protein encoded by the gene		
CC	clone detailed in the descriptor line. The gene can be used to generate		
CC	fusion proteins by linking to the gene to a human immunoglobulin Fc		
CC	portion (e.g. AAZ06210) for increasing the stability of the fused		
CC	protein as compared to the human protein only.		
CC	The invention relates to 36 novel genes and their fragments (nucleic		
CC	acid sequences; AAZ06219-206263; amino acid sequences AAY38386-Y38498)		
CC	which are useful for preventing, treating or ameliorating medical		
CC	conditions e.g. by protein or gene therapy. Also, pathological		
CC	conditions can be diagnosed by determining the amount of the new		
CC	polypeptides in a sample or by determining the presence of mutations in		
CC	the new polynucleotides. Specific uses are described for each of the 36		
CC	polynucleotides, based on which tissues they are most highly expressed in		
CC	(see AAZ06219 for described uses).		
XX			
SQ	Sequence 98 AA;		
Alignment Scores:			
Pred. No.:	1.53e-27	Length:	98
Score:	417.00	Matches:	88
Percent Similarity:	83.81%	Conservative:	0

XX OS Homo sapiens.
XX PN WO200207751-A1.
XX PD 31-JAN-2002.
XX PF 24-JUL-2001; 2001WO-US23269.
XX PR 25-JUL-2000; 2000US-221215P.
XX PR 28-NOV-2000; 2000US-0723810.
XX PA (AXCE-) AXCELL BIOSCIENCES CORP.
XX PI Herrero J, Pirozzi G, Uveges A;
XX DR N-PSDB; ABK46540.
XX DR WPI; 2002-195842/25.
XX DR N-PSDB; ABK46540.
XX PT Methods for identifying polypeptides comprising PDZ domains, the
XX PT polypeptides and their encoding nucleic acids, useful for the diagnosis
XX PT and treatment of PDZ related disorders -
XX PS Claim 43; Fig 18; 225pp; English.
XX CC The invention relates to methods for identifying polypeptides comprising
XX CC PDZ domains, and their encoding nucleic acids. The sequences are used to
XX CC identify modulators of their expression, function and activity, for use
XX CC in the diagnosis and treatment of PDZ related disorders. Antibodies
XX CC against the proteins and cells that produce them may be used for the
XX CC treatment of PDZ-mediated disease states. Sequences AAU87843-AAU87974
XX CC represent proteins containing PDZ domains, fragments of these proteins
XX CC and other related peptides used in the methods of the invention.
XX SQ Sequence 1535 AA;

Alignment Scores:
Pred. No.: 2,56e-14 Length: 1535
Score: 271.00 Matches: 177
Percent Similarity: 32.17% Conservative: 91
Best Local Similarity: 21.25% Mismatches: 252
Query Match: 6.79% Indels: 315
DB: Gaps: 31

US-09-502-945-5 (1-2162) x AAU87934 (1-1535)

Qy 289 CTGATCCCACTGAACACACAGTGGAAATATGATCATGCTGACCCCGCGCTCCAGGAAG 348
Db 744 IleThrProLeuArgProGluAspAspTyrSerPro-----ArgGluArgArg 759
Qy 349 CTGAAGGAGGTGCGTCTGACCGTGTGACCCCGAA----- 384
Db 760 GlyGlyLeuArgLeuProLeuLeuProGlyProGlyProGlyProLeuArgGlnArg 779
Qy 385 -----GGCTTCGCGCTGCTCATCTCCACCTCATCAAGGC 411
Db 780 HisValAlaCysLeuAlaArgSerGluArgGlyLeuGlyPheSerIleAlaGlyLys 799
Qy 412 -----CTGGAGTTTGGCTGTGGCTTCTCATCTCCACCTCATCAAGGC 456
Db 800 GlySerThrProTyrArgAlaGlyAspAlaGlyIlePheValSerArgIleAlaGly 819
Qy 457 GGTACGGACAGACGCTCGGG---CTCCAGGTAGGGACGAGATCGTCCGGATCAATGGA 513
Db 820 GlyAlaAlaHisArgAlaGlyThrLeuGlnValGlyAspArgValLeuSerIleAsnGly 839
Qy 514 TATTCATCTCTCTCTACCATGAGGAGGTCTACACCTCAT---CGAACCAAGAAA 570
Db 840 ValAspValThrGluAlaArgHisAlaValSerLeuLeuThrAlaAlaSerPro 859
Qy 571 ACTGTGTGCATCAAAAGTGAGACACATCGCGCTGATCCCGGTGAAAAGCTCTCCT----- 624
Db 860 ThrIleAlaLeuLeuLeuGluArgGluAlaGlyProLeuProProSerProLeuPro 879

Qy 625 -----GATGAGCCCTCACTTGGCAGTATGTCGATCAGTTTGTGCGAATCTGGGGC 678
Db 880 HisSerSerProThrAlaAlaValAlaThrThrSerIleThrThrAlaThrProGly 899
Qy 679 GTGCGAGC----- 687
Db 900 ValProGlyLeuProSerLeuAlaProSerLeuLeuAlaAlaLeuGluGlyProTyr 919
Qy 688 -----AGCTGGCTCCCTGGAAATCGGAAACAAAGAG 723
Db 920 ProValGluGluIleArgLeuProArgAlaGlyPro----- 932
Qy 724 AAGAGGTCTTCATCAGCCTGGTGGCTCCGAGCGCTTGGCTGCAGCATTTCCAGCGC 783
Db 933 -----LeuGlyLeuSerIleValGly-----GlySerAspHisSerHis 946
Qy 784 CCC-----ATCCAGAACCTGGCATCTTTATCAGCCATGTGAACCTGGCTCCTGCT 837
Db 947 ProPheGlyValGlnGluProGlyValPheIleSerLysValLeuProArgGlyLeuAla 966
Qy 838 GCTGAGTGGGATGAGATAGGGACGAGATTGTGGAAGTCAATGGCGTGCAGCTTCTCT 897
Db 967 AlaArgSerGlyLeuArgValGlyAspArgIleLeuAlaValAsnGlyGlnAspValArg 986
Qy 898 AACCTGATCATCAGAGGCTGTAAATGTGCTG----- 930
Db 987 AspAlaThrHisGlnGluAlaValSerAlaLeuLeuArgProCysLeuGluLeuSerLeu 1006
Qy 930 ----- 930
Db 1007 LeuValArgArgAspProAlaProProGlyLeuArgGluLeuCysIleGlnLysAlaPro 1026
Qy 931 -----AAAAATAGCCG 942
Db 1027 GlyGluArgLeuGlyIleSerIleArgGlyAlaArgGlyHisAlaGlyAsnProArg 1046
Qy 943 AGCTGAGC-----ATCTCCATTGTAGCTGCGAGCTGGCGCG 978
Db 1047 AspProThrArgGluGlyIlePheIleSerLysValSerProThrGlyAlaAlaGlyArg 1066
Qy 979 GAGCTGTTTCATGACAGACCGGCGCTGGCAGCGCGCGGCGAGCTGAGCTG----- 1032
Db 1067 AspGlyArgLeuArgValGlyLeuArgLeuGluValAsnGlnSerLeuLeuGly 1086
Qy 1033 -----CAGCGCAGAGCTTCTCATGCA----- 1055
Db 1087 LeuThrHisGlyGluAlaValGlnLeuLeuArgSerValGlyAspThrLeuThrValLeu 1106
Qy 1056 -----GAAGCGGTGGCGATGGAGTCCCAACAGATCCTCCAGGA----- 1094
Db 1107 ValCysAspGlyPheGluAlaSerThrAspAlaAlaLeuGluValSerProGlyValIle 1126
Qy 1095 -----GCAGCAGGAGATGGA-----GCGGCAAG 1118
Db 1127 AlaAsnProPheAlaAlaGlyIleGlyHisArgAsnSerLeuGluSerIleSerIle 1146
Qy 1119 GAGAAAGAAATTCGCCAGAGGCGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1178
Db 1147 AspArgGluLeuSerProGluGlyProGlyLysGluLysGluLeuPro-----Gly 1163
Qy 1179 ACAGATTGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1223
Db 1164 GlnThrLeuHisTrpGlyProGluAlaThrGluAlaAlaGlyArgGlyLeuGlnProLeu 1183
Qy 1224 AGACTGGGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1283
Db 1184 LysLeuAspTyrArgAlaLeuAlaAlaVal-----Pro 1194
Qy 1284 AGTACCCCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1343
Db 1195 SerAlaGlySerValGlnArgValProSerGlyAlaAlaGlyGlyLysMetAlaGluSer 1214


```
QY 1344 CTATGTTCCACCCAGAGCAGATCATGGGAGGATGTCGGGCTCTACGCATCAAGAAGG 1403
Db 1215 ProCysSerProSerGlyGln-GlnProProSerProSerProAspGluLeuProAl 1234
QY 1404 AGGGATCCTTAGA----- 1416
Db 1234 aAsnValLysGlnAlaTyArgAlaPheAlaValProThrSerHisProProGluAs 1254
QY 1417 -----CCTGGCCCTCGAAG----- 1431
Db 1254 pAlaProAlaGlnProProThrProGlyProAlaAlaSerProGluGlnLeuSerPheAr 1274
QY 1432 -----CGGTGTGGATCCTCCCATTTGGGAAGT-----GG 1460
Db 1274 gGluArgGlnLysTyPheGluLeuGluValArgValProGlnAlaGluGlyProProLy 1294
QY 1461 TCGTTTCTGCTGTATGAGCGGGAGCTGCTGACGGCATGGTGGCATTTGTGAAGGGG 1520
Db 1294 sArgValSerLeuVal-----G 1300
QY 1521 ACGAGATCATGCGCAATACAGCGCAAGATTGTGACAGACTACACCCTGCTGAGGCTGACG 1580
Db 1300 yAlaAspAspLeuArgLysMetGlnGlu----- 1309
QY 1581 CTGCCCCGAGAAGGCCTTGGATCAGCGGGGACTGGATCGACCTTGTGTGCGCTCT 1640
Db 1310 -----GluGluAlaArgLysLeuGlnGln----- 1317
QY 1641 GCCCCCCAAGGAGTATGACGATGAGCTGCTTCTGTGAGTCCAAAGGGGAACCC 1700
Db 1318 -----LysArgAlaGlnMetLeuArgGluAlaAlaGluAlaGlyAl 1331
QY 1701 AAATTCACGCGTTAGGAACAGTGTGAGCTCCGCGCCACCTGTGAACACAAAGCCTCGGA 1760
Db 1331 aGluAlaArgLeuAlaLeuAspGlyGluThrLeuGluGluGluGlnGluAspGluCl 1351
QY 1761 CCAGCCTTGAGAGGCCATGATGACACACAGATGCGATCCTTGGGACCTGAATCTAT 1820
Db 1351 nProProTAlaSerProSerProThrSerArgGlnSerProAlaSerProProProLe 1371
QY 1821 C-----ACCCAGGAATCTCAACTCC 1841
Db 1371 uGlyGlyAlaProValArgThrAlaLysAlaGluArgHisGlnGluArgLeuArgVa 1391
QY 1842 CTTTGGCCTTGACACAGGCGCAGATAGGAACAGCTCGGCGCCACTTTT-TTGAAGGCCAA 1900
Db 1391 lGlnSerProGluProProAlaProGluArgAlaLeuSerProAlaGluLeuArgAla-- 1410
QY 1901 TGTGGAGAAAGGAGCAGCAGCCGTTTGGGAGAAGATCTCAAGATCCAGACTCTCAT 1960
Db 1411 -----LeuGluAlaGluLysArgAlaLeuTrp---ArgAlaAlaArgMetLysSerLeuGl 1428
QY 1961 TCCTTTCCTCTGCGCCAGTGAATTTGGTCTCTCCAGCTTTGG-----GGGACT-- 2009
Db 1428 uGlnAspAlaLeuArgAlaGlnMetValLeuSerArgSerGlnGluGlyArgGlyThrAr 1448
QY 2010 -----CCTTCCTTGAACCCCTAATAAGACCCC 2035
Db 1448 gGlyProLeuGluArgLeuAlaGluAlaProSerProAlaProThrProSerProThrPr 1468
QY 2036 ACTGGAGTCTCTC----- 2048
Db 1468 oValGluAspLeuGlyProGlnThrSerThrSerProGlyArgLeuSerProAspPheAl 1488
QY 2049 -----TCTCTCCATCCCTCTCTCTGCCC 2072
Db 1488 aGluGluLeuArgSerLeuGluProSerProSerPro 1500
RESULT 13
ABG22366
ID ABG22366 standard; Protein; 1037 AA.
XX
AC ABG22366;
```

```
XX 18-FEB-2002 (first entry)
DT
DE
XE
XX Novel human diagnostic protein #22357.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS86553.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
FS Claim 20; SEQ ID No 52725; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1037 AA;
Alignment Scores:
Pred. No.: 4,28e-14 Length: 1037
Score: 267.50 Matches: 157
Percent Similarity: 35.09% Conservativeness: 89
Best Local Similarity: 22.40% Mismatches: 220
Query Match: 6.71% Indels: 235
DB: Gaps: 32
US-09-502-945-5 (1-2162) x ABG22366 (1-1037)
QY 183 GCTGGGAATGTACACACAGAC-----CATGACGTCGGCGCTGCTGGGAGACT 233
Db 83 SerAlaIleValProGlyAspGlyValAlaGlyGlyGlyThrGlyAspGlyGlyPro 102
QY 234 GAAGCTGTTCATCAATGAACCCAGCCGCTGCTCTGTTTGTATGCCATTCGCCCTGAT 293
III ;;
```


Db 103 GlyAlaProTyr----- 106
QY 294 CCCACTGAGCACCAGCTGGATATGATCAGCTGACCGCCCGCCCTCCAGGAACCTGAA 353
Db 107 -----ProGlyAlaValAlaAspMetGluGluLeuThrIleTrpGluGlnHis 122
QY 354 GGAGGTGGCTCTGGACCGCTCTCACCCGCAAGCGCTCGCCTGAGCTGGCTGGCGCT 413
Db 123 ThrAlaThrLeuSerLysAspProArgArg-GlyPheGlyIleAlaIleSerGlyGlyArg 142
QY 414 GGAGTTT---GGCTGTGGCTCTTCATCTCCACCTCATCAAGCGCGTACGAGCAGACAG 470
Db 142 gAspArgProGlyGlySerMetValValSerAspValProGlyGlyProAlaGluG1 162
QY 471 CGTCGGGTCCAGGTAGGGAGCAGATCGTCGGGATCAATGGATATTCATCTCCTCTG 530
Db 162 y---ArgLeuGlnThrGlyAspHisIleValMetValAsnGlyValSerMetGluAsnAl 181
QY 531 TACCCATGAGGAGGTCAACCTCATTCGAACC---AAGAAACTGTGTCCATCAAAAGT 587
Db 181 aThrSerAlaPheAlaIleGlnIleLeuLysThrCysThrLysMetAlaAsnIleThrVa 201
QY 588 G-----AGACATATCGGCTGATCCCGTGAAGAGCTCTCT----- 624
Db 201 LlysArgProArgIleHisLeuProAlaThrLysAlaSerProSerSerProGlyArg 221
QY 625 -----GATGACCCCTCACTTGGCGATATGTGGATCAG----- 657
Db 221 gGlnAspSerAspGluAspAspGlyProGlnArgValGluGluValAspGlnGlyArgG1 241
QY 658 -----TTTGTGTGGATCTGG----- 675
Db 241 yTyraSpGlyAspSerSerSerGlySerGlyArgSerTrpAspGluArgSerArgArgPr 261
QY 676 -----GGCTGCGAGGAGCGCTGGCTCCCTCGAAATCGGAAACAAAGGAGAGAA 728
Db 261 oArgProGlyArgArgGlyArgAlaGlySerHisGlyArgSerProGlyGly----- 279
QY 729 GGTCTTCATCAGCTGTAGGCTCCGAGGCTGGCTGGCAGCATTTCCAGCGG----- 783
Db 280 -----GlySerGluAlaAsnGlyLeuAlaLeuValSerGlyPheLy 293
QY 784 -----OCCATCCAG-----AAGCCT----- 798
Db 293 sArgLeuProArgGlnAspValGlnMetLysProValLysSerValLeuValLysArgArg 313
QY 799 -----GGCATCTTTATCAGCCATGTGAAC 824
Db 313 gAspSerGluGluPheGlyValLysLeuGlySerGlnIlePheIleLysHisIleThrAs 333
QY 825 TGGCTCCCTGTCTGCT---GAGGTGGGATTGGAGATAGGGACAGATTGTGCAAGTCAA 881
Db 333 pSerGlyLeuAlaAlaArgHisArgGlyLeuGlnGluGlyAspLeuIleLeuGlnIleAs 353
QY 882 TGGCTGACCTCTCTAACCCTGGATCACAAGAGGCTGTAATGTGCTGAAATATAGCCG 941
Db 353 nGlyValSerSerGlnAsnLeuSerLeuAsnAspThrArgLeuIleGluLysSerGl 373
QY 942 CAGCCTGACCATCTCATTGTAGCTGACGCTGGCGGAGCTGTTCATGACAGAC----- 996
Db 373 uGly---LysLeuSerLeuValLeuArgAspArgGlyGlnPheLeuValAsnIlePr 392
QY 997 -----CGGAGCG 1004
Db 392 oProAlaValSerAspSerSerProLeuGluAspIleSerAspLeuAlaSerGl 412
QY 1005 GCTGCAGAGCG-----CGGAGCGTGGCTGAGCTGAGCGCGCAGGA 1043
Db 412 uLeuSerGlnAlaProProSerHisIleProProProProArgHisAlaGlnArgSerPr 432
QY 1044 GCTTCTCATGCAAGCGCTGGCGATGTCACCAAGATCTCTCCAGGAGCAGCAGGA 1103
Db 432 oGluAlaSerGlnThrAspSerProValGluSerProArgLeu----- 446

QY 1104 GATGAGCGGCAAGAGGAGAAAGAAATTTGCCAGAAAGCAGCAGAGGAAATAGAGATA 1163
Db 447 -----ArgArgGluSerSerValAspSerArgThrIleSerGluProAspGluG1 463
QY 1164 CCGAAGAGAGATGAACAGATTGTAGAGGAGGAGAAAGATTTTAAAGCAATCGGAAGA 1223
Db 463 nArgSerGlu----- 466
QY 1224 AGACTGGGCTCAAAGGAACAGCTACTCTTGCTTAAACCATCATCTGCTGAGGTACACCC 1283
Db 467 -----LeuProArgGluSerSerTyraSpIleTyra 477
QY 1284 AGTACCCCTTCGCAAGCAAAAGTATTTCCGGAATATAGAGAAAGGCTTTGACCCCTACT 1343
Db 477 gVal-Pro---SerSerGlnSer-----MetGluAspArgGlyTyraSerPro--- 491
QY 1344 CTATGTTTCAACCCAGACAGCATCATGGGAAGGATGTCCGGCTCTACGCATCAAGAGG 1403
Db 492 -----AspThrArgValValArgPheLeuLysG 501
QY 1404 AGGATCTTAGACCTGGCCCTGGAAGCGGTGTGGACTCCCCCATTTGGAAAGTGTGTCG 1463
Db 501 LysSerIleGlyLeuArgLeuAlaGlyGlyAsnAsp-----ValGly---IlePheV 518
QY 1464 TTTCTGTGTATAGCGGGGAGCTCTGAGCGGCATGGTGGCATTTGTGAAGGGGAGC 1523
Db 518 aLysGlyVal---GlnAlaGlySerProAlaAspGlyGlnGlyIleGlnGluGlyAspG 537
QY 1524 AGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTGGCTGGCTGAGCGCTG 1583
Db 537 InIleLeuGlnValAsnAspValProPheGlnAsnLeuThrArgGluGluAla----- 554
QY 1584 CCCTGCAGAGGCTCGAATCAGGGCGGAGCTGGATCGACCTGTGTGGTTCGCTGTCGC 1643
Db 555 -----ValGlnPheLeuLeuGlyLeuProp 563
QY 1644 CCCCAAAGAGATATGAC-----GATGAGCTGACCTTCTTCTGCTGA 1682
Db 563 roGlyGluGluMetGluLeuValThrGlnArgLysGlnAspIlePheTrpLysMetValG 583
QY 1683 AGTCCAAAGGGGAAACAAATTCACCGCTTAGGAACACAGTACGCTCCGCGCCACCTCG 1742
Db 583 InSerArgValGlyAspSerPheTyraIleArgThrHisPheGluLeuGluPro----- 600
QY 1743 TGAACAAAGCTCTCGGACCGCTT-----GAGAGAGCGCCACATGACACACCC----- 1792
Db 601 -----SerProProSerGlyLeuGlyPheThrArgGlyAspValPheHisValLeuA 618
QY 1793 --AGATGGCATCTTGGGACCTGAATCTATCATCCAGGAATCTCAAACTCCCTTTGGCCC 1850
Db 618 spThrLeuHisPro----- 622
QY 1851 TGAACAGGCGCCAGATAAGGNACAGCTCGGCGCACCTTTTGAAGGCCAATGTGGAGGAA 1910
Db 623 --GlyProGlyGlnSerHisAlaArgGlyGlyHisTrpLeu----- 635
QY 1911 AGGAGCAGCAGCGCTTTGGGAGAAATCTCAAGGATCCAGACTCTCATTTCTTCTTCT 1969
Db 636 -----AlaValArgMetGlyArgAspLeuArgGluGlnGluArgGlyIleIlePro 652
RESULT 14
AAE21718
ID AAE21718 standard; Protein; 928 AA.
XX AC AAE21718;
XX AC
DT 16-JUL-2002 (first entry)
XX DE Human PKIN-13 protein.
XX DE
KW Human; kinase; enzyme; PKIN-13 protein; immune system disorder; anaemia;
KW acquired immune deficiency syndrome; thymic hypoplasia; Crohn's disease;

KW asthma; neurological disorder; epilepsy; Charcot-Marie-Tooth disease;
 KW AIDS; seizures; cell proliferative disorder; cancer; adenocarcinoma;
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; developmental disorder;
 KW Down's syndrome; gene therapy; protein therapy; cytostatic.

XX Homo sapiens.

XX Key Location/Qualifiers
 XX Domain 10..99 /note= "GLGF domain"
 XX Domain 20..101 /note= "PDZ domain"
 XX Domain 204..280 /note= "PDZ domain"
 XX Domain 388..469 /note= "GLGF domain"
 XX Domain 391..471 /note= "PDZ domain"
 XX Domain 429..439 /note= "PDZ domain"
 XX Domain 733..754 /note= "Leucine zipper domain"

XX WO200218557-A2.

XX PD 07-MAR-2002.

XX 31-AUG-2001; 2001WO-US27219.

XX 31-AUG-2000; 2000US-229873P.

XX 08-SEP-2000; 2000US-231357P.

XX 14-SEP-2000; 2000US-232654P.

XX 22-SEP-2000; 2000US-234902P.

XX 29-SEP-2000; 2000US-236499P.

XX 06-OCT-2000; 2000US-238389P.

XX 13-OCT-2000; 2000US-240542P.

XX (INCY-) INCYTE GENOMICS INC.

XX Bandman O, Nguyen DB, Wallia NK, Hafalia AJA, Yao MG, Gandhi AR;

XX Gururajan R, Ding L, Patterson C, Yue H, Baughn MR, Tribouley CM;

XX Thornton M, Elliott VS, Lu Y, Ison CH, Au-Young J, Tang YF;

XX Azimzai Y, Burrill JD, Marcus GA, Zingler KA, Lu DAM, Lal PG;

XX Ramkumar J, Warren BA, Kearney L, Policky JL, Thangavelu K;

XX Burford N;

XX WPI; 2002-329769/36.

XX N-PSDB; AAD34310.

XX New human kinases, useful for diagnosing, treating or preventing immune system disorders (e.g. Crohn's disease), neurological disorders (e.g. epilepsy), or cell proliferative disorders (e.g. cancers such as leukemia or lymphoma)

XX Claim 68; Page 174-176; 218pp; English.

XX The present invention relates to human kinases (PKIN) and polynucleotides encoding such proteins. PKIN sequences of the invention are useful for diagnosing, treating or preventing disorders associated with aberrant expression of PKIN, particularly immune system disorders (e.g. acquired immune deficiency syndrome (AIDS), thymic hypoplasia, Crohn's disease, anaemia, asthma), neurological disorders (e.g. epilepsy, Charcot-Marie-Tooth disease or seizures), cell proliferative disorders (e.g. cancers such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma), and developmental disorders (e.g. Down's syndrome). They are also used in gene therapy and protein therapy. The present sequence is human PKIN-13 protein.

XX Sequence 928 AA;

XX Alignment Scores:

XX Pred. No.: 4.5e-14 Length: 928

XX Score: 267.00 Matches: 144

Percent Similarity: 36.19% Conservative: 84
 Best Local Similarity: 22.86% Mismatches: 191
 Query Match: 6.69% Indels: 211
 DB: 23 Gaps: 30
 US-09-502-945-5 (1-2162) x AAE21718 (1-928)
 QY 385 GGCTCGGCTGAGTGTGGTGGTGGCTGGAGTTT---GGTGTGGGCTCTTCATCTCC 441
 Db 30 GlyPheGlyIleAlaIleSerGlyArgAspArgProGlyGlySerMetValSer 49
 QY 442 CACCTCATCAAGCGGTGAGGACAGCGTGGCTCCAGGTAGGACGATCGTC 501
 Db 50 AspValValProGlyGlyProAlaGluGly---ArgLeuGlnThrGlyAspHisIleVal 68
 QY 502 CGGATCAATGGATATTCATCTCTCTGTACCCATGAGGAGTCATCAACCTCATCGA 561
 Db 69 MetValAsnGlyValSerMetGluAsnAlaThrSerAlaPheAlaIleGlnIleLeuLys 88
 QY 562 ACC---AAGAAACTGTGCTCCATCAAGTG-----AGACACATCGGCTGATCCCC 609
 Db 89 ThrCysThrLysMetAlaAsnIleThrValLysArgProArgIleHisLeuProAla 108
 QY 610 GTGAAAGCTCTCT-----GATGAGCCCTC 636
 Db 109 ThrLysAlaSerProSerSerProGlyArgGlnAspSerAspGluAspAspGlyProGln 128
 QY 637 ACTTGGCAGTATGTGGATCAG-----TTTGTGTCGGAATCTGGG 675
 Db 129 ArgValGluGluValAspGlnGlyArgGlyTyrAspGlyAspSerSerSerGlySerGly 148
 QY 676 -----GGCTGGGAGGACGCTGGGCTCC 699
 Db 149 ArgSerTrpAspGluArgSerArgArgProArgProGlyArgGlyArgAlaGlySer 168
 QY 700 CCTGGAATCGGAAACAAGGAGAAGGTCTTCATCAGCTGTGTAGCTCCGAGGC 759
 Db 169 HisGlyArgArgSerProGlyGly-----GlySerGluAla 180
 QY 760 CTGCTGCTGACATTTCCAGCGC-----CCATCCAG-----AAG 795
 Db 181 AsnGlyLeuAlaLeuValSerGlyPheLysArgLeuProArgGlnAspValGlnMetLys 200
 QY 796 CCT----- 798
 Db 201 ProValLysSerValLeuValLysArgAspSerGluGluPheGlyValLysLeuGly 220
 QY 799 --GGCATCTTTATCAGCCATGTGAAACCTGGCTCCTGTCTGCT---GAGGTGGGATTG 852
 Db 221 SerGlnIlePheIleLysHisIleThrAspSerGlyLeuAlaAlaArgHisArgGlyLeu 240
 QY 853 GAGATAGGGACCAAGATTGTGAAAGTCAATGGCTCGACTTCTTAACCTGGATCACAAG 912
 Db 241 GlnGluGlyAspLeuIleLeuGlnIleAsnGlyValSerSerGlnAsnLeuSerLeuAsn 260
 QY 913 GAGCTGTAAATGTGCTGAAAATAGCGGAGCTGACCATCTCCATGTAGCTGCAGCT 972
 Db 261 AspThrArgArgLeuIleGluLysSerGluGly---LysLeuSerLeuLeuValLeuArg 279
 QY 973 GGCCGGGAGCTGTTTCAGCAGAC----- 996
 Db 280 AspArgGlyGlnPheLeuValAsnIleProProAlaValSerAspSerAspSerPro 299
 QY 997 -----CGGAGCGGCTGCAGAGCG----- 1017
 Db 300 LeuGluAspIleSerAspLeuAlaSerGluLeuSerGlnAlaProSerHisIlePro 319
 QY 1018 ---CGGCAGCTGAGCTGCGGAGAGCTTCTCATGAGAGCGGCTGGCGATGAG 1074
 Db 320 ProProArgHisAlaGlnArgSerProGluAlaSerGlnThrAspSerProValGlu 339
 QY 1075 TCCACAGATCTCCAGGAGCAGCAGAGATGAGCGGCAAGAGAAAGAAATGCC 1134
 Db 1075 TCCACAGATCTCCAGGAGCAGCAGAGATGAGCGGCAAGAGAAAGAAATGCC 1134

Dh 340 SerProArgLeu-----ArgArgGluSerSerValAsp 350
Qy 1135 CAGAGGCGAGAGAGAAAATGAGAGATACCGGAGAGATGGAAACAGATTGTAGAGGAG 1194
Dh 351 SerArgThrIleSerGluProAspGluArgSerGlu----- 363
Qy 1195 GAAGAGAAGTTAAGAACCATGGGAAGAACTGGGCTCAAGGAAACACACTACTCTTG 1254
Dh 364 -----Leu 364
Qy 1255 CTAAGACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATTTCG 1314
Dh 365 ProArgGluSerSerTyrAspIleTyrArgVal-Pro---SerSerGlnSer----- 380
Qy 1315 GAAATATGAGGAGGCTTTGACCCCTACTCTATGTTACCCAGCAGACATCATGGGAA 1374
Dh 381 -MetGluAspArgGlyTyrSerPro----- 388
Qy 1375 GGATGTCGGCTCTACGATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAGGCGG 1434
Dh 389 -AspThrArgValValArgPheLeuIleGlySerIleGlyLeuArgLeuAlaGlyG 408
Qy 1435 TGTGGACTCCCCCATGTTGGAGAGGTGCTTCTGCTGTATGAGCGGGAGCTGCTGA 1494
Dh 408 yAsnAsp-----ValGly---IlePheValSerGlyVal---GlnAlaGlySerProAl 424
Qy 1495 CGGCATGCTGCATTTGAAGGGGAGGAGATCATGCAATCAACGGCAAGATGTGAC 1554
Dh 424 aAspGlyGlnGlyIleGlnGlyArgPheGlnIleLeuGlnValAsnAspValProPheG 444
Qy 1555 AGACTACACCTGGCTGAGGTGAGCTGCGCTCGCAAGGCGCTGGAATCAGGCGGGGA 1614
Dh 444 nAsnLeuThrArgGluGluAla----- 451
Qy 1615 CTGATGACGACCTGTGTGTCGCTGCCGCCCAAGAGGATGATGAC----- 1660
Dh 452 ---ValGlnPheLeuLeuGlyLeuProProGlyGluGluMetGluLeuValThrGlnAr 470
Qy 1661 -----GATGAGCTGACCTTCTGCTGAAGTCCAAAGGGAACCAATTCACGGGTT 1713
Dh 470 yGlyGlnAspIlePheThrPylsMetValGlnSerArgValGlyAspSerPheTyrIleAr 490
Qy 1714 AGAAACAGTGCAGCTCGCGGCCACCTCGTGAACACAAAGCTCGGACACGCTT----- 1768
Dh 490 gThrHisPheGluLeuGluPro-----SerProSerGlyLeuGlyPh 505
Qy 1769 -GAGAGGCGCACATGACACACACC-----AGATGGCATCCTTGGGACCTGAATCTATC 1821
Dh 505 eThrArgGlyAspValPheHisValLeuAspThrLeuHisPro----- 519
Qy 1822 ACCAGGATCTCAACTCCCTTTGGCCCTGAACGAGGCGCAGATAGGAACACGCTCGG 1881
Dh 520 -----GlyProGlyGlnSerHisAlaAArgGlyG 529
Qy 1882 CCACCTTTTGAAGGCCAATGTGGAGGAAGGAGCAGCGCTTGGGAGAGATCT 1941
Dh 529 yHisTrpLeu-----AlaValArgMetGlyArgAspLe 540
Qy 1942 CAAGATCCAGACTCTCATCTTCTTCCT 1969
Dh 540 uArgGluGlnGluArgGlyIleIlePro 549
RESULT 15
AAB01383
ID AAB01383 standard; Protein; 856 AA.
XX
AC AAB01383;
XX
DT 20-OCT-2000 (first entry)
XX
DE Neuron-associated protein.
XX
KW Neuron associated protein; NEUAP; neurological disorder; epilepsy;

ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW Alzheimer's disease; Pick's disease; Huntington's disease;
KW dementia; Parkinson's disease; demyelinating disease; meningitis;
KW prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;
KW cerebral palsy; muscular dystrophy; central nervous system; CNS;
KW peripheral nervous system; PNS; myopathy; schizophrenia;
KW actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW AIDS; Addison's disease; inflammation; acquired immunodeficiency syndrome;
KW allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW werner syndrome; trauma; human.
XX Homo sapiens.
XX OS
FH Domain
FT 1..65
FT /label= PD2 domain signature
FT Modified-site 18
FT /note= "Potential glycosylation site"
FT Modified-site 20
FT /note= "Potential phosphorylation site"
FT Modified-site 69
FT /note= "Potential phosphorylation site"
FT Modified-site 71
FT /note= "Potential phosphorylation site"
FT Modified-site 73
FT /note= "Potential phosphorylation site"
FT Modified-site 118
FT /note= "Potential phosphorylation site"
FT Modified-site 119
FT /note= "Potential phosphorylation site"
FT Domain 136..218
FT /label= PD2 domain signature
FT Modified-site 144..147
FT /note= "Glycosaminoglycan attachment site"
FT Modified-site 199
FT /note= "Potential glycosylation site"
FT Modified-site 238
FT /note= "Potential phosphorylation site"
FT Modified-site 253
FT /note= "Potential phosphorylation site"
FT Modified-site 263
FT /note= "Potential phosphorylation site"
FT Domain 269..349
FT /label= PD2 domain signature
FT Modified-site 284
FT /note= "Potential phosphorylation site"
FT Modified-site 296
FT /note= "Potential phosphorylation site"
FT Modified-site 337
FT /note= "Potential phosphorylation site"
FT Modified-site 341
FT /note= "Potential phosphorylation site"
FT Modified-site 369
FT /note= "Potential glycosylation site"
FT Modified-site 370
FT /note= "Potential phosphorylation site"
FT Modified-site 386
FT /note= "Potential phosphorylation site"
FT Modified-site 389
FT /note= "Potential glycosylation site"
FT Modified-site 390
FT /note= "Potential phosphorylation site"
FT Modified-site 414
FT /note= "Potential phosphorylation site"
FT Domain 415..497
FT /label= PD2 domain signature
FT Modified-site 475
FT /note= "Potential phosphorylation site"
FT Domain 511..592
FT /label= PD2 domain signature

FT Modified-site 531 /note= "Potential glycosylation site"
 FT Modified-site 568 /note= "Potential glycosylation site"
 FT Modified-site 599 /note= "Potential glycosylation site"
 FT Modified-site 604 /note= "Potential phosphorylation site"
 FT Modified-site 625 /note= "Potential phosphorylation site"
 FT Modified-site 633 /note= "Potential phosphorylation site"
 FT Modified-site 634 /note= "Potential phosphorylation site"
 FT Domain 648..733 /label= "Potential phosphorylation site"
 FT Modified-site 690 /label= PDZ domain signature
 FT Modified-site 705 /note= "Potential phosphorylation site"
 FT Modified-site 721 /note= "Potential phosphorylation site"
 FT Domain 773..856 /note= "Potential glycosylation site"
 FT Region 819..821 /label= PDZ domain signature
 FT Modified-site 835 /label= Cell attachment sequence
 FT Modified-site /note= "Potential phosphorylation site"
 XX WO200034477-A2.
 PX 15-JUN-2000.
 PD 10-DEC-1999; 99WO-US30408.
 PF 11-DEC-1999; 98US-0210083.
 PR 09-FEB-1999; 98US-9123456.
 PR 16-MAR-1999; 99US-0119365.
 XX (INCY-) INCYTE PHARM INC.
 XX Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-young J, Yang J;
 PI Lu DAM, Azimzai Y;
 XX WPI: 2000-423423/36.
 DR N-PSDB; AAA47424.

XX New human neuron-associated proteins and polynucleotides encoding them,
 PT useful for diagnosis, treatment and prevention of cell proliferative
 PT disorders including cancer, neuronal and neurological disorders
 XX Claim 1; Page 105-108; 145pp; English.
 PS Human neuron-associated proteins (NEUPAP) can be used for for
 CC treating or preventing a disorder associated with decreased
 CC expression or activity of NEUPAP. Antagonists of NEUPAP are useful for
 CC treating or preventing disorder associated with increased expression
 CC or activity of NEUPAP. NEUPAP or their fragments or derivatives are
 CC useful for treating neurological disorder such as epilepsy, ischemic
 CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
 CC disease, Pick's disease, Huntington's disease, dementia and
 CC Parkinson's disease. NEUPAPs are also useful for treating other
 CC demyelinating diseases, bacterial and viral meningitis, prion
 CC diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
 CC metabolic diseases of the nervous system, neurofibromatosis, other
 CC developmental disorders of the central nervous system, cerebral
 CC palsy, neuroskeletal disorders, autonomic nervous system disorders,
 CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and
 CC other neuromuscular disorders, peripheral nervous system disorders,
 CC inherited, metabolic, endocrine, and toxic myopathies, mental
 CC disorders including mood, anxiety and schizophrenic disorders, a cell
 CC proliferative disorder such as actinic keratosis, arteriosclerosis,

CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
 CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
 CC haemoglobinuria, cancers of the adrenal gland, bladder, bone,
 CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
 CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
 CC disease, adult respiratory distress syndrome, allergies, ankylosing
 CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
 CC complications of cancer, hemodialysis, and extracorporeal circulation,
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic
 CC infections, and trauma. This protein was given the incyte ID no.
 CC 3201753CD1.

XX Sequence 856 AA;

SQ Alignment Scores:

Pred. No.: 9.63e-14 Length: 856
 Score: 263.00 Matches: 146
 Percent Similarity: 36.47% Conservative: 75
 Best Local Similarity: 24.09% Mismatches: 178
 Query Match: 6.59% Indels: 207
 DB: 21 Gaps: 26

US-09-502-945-5 (1-2162) x AAB01383 (1-856)

QY 11 TCGCGTGGCGGCTCTTCCAGTCTCTGGCAGCGCGGCCGAA-----GGAACG 61
 |||:||||| :||| ||||| ||||| |||||
 Db 371 SerGlnSerLeuMetValProGlnSerGlySerProGluProGluSerIleArgAsnThr 390
 QY 62 GGTGCTCAACGACGACCTGGACCTGCCCGACCATGGACCAAGTGGCCGAGCAAT 121
 |||:||||| :||| ||||| ||||| |||||
 Db 391 SerArgSerThrProAlaIlePheAla-Ser--AspProAlaThrCysProIleI 409
 QY 122 TCCGCGCATAGGTGGATTTCTGATTGAAATGATGC----- 158
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 409 eProGly-----CysGluThrThrIleGluIleSerLy 420
 QY 159 ----AGAGAGGACTACTCTATGATGCTGCTG----- 186
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 420 sGlyArgThrGlyLeuGlyLeu-SerIleValGlyGlySerAspThrLeuLeuGlyAlaI 440
 QY 187 -----CGAATGTACCACAG-----ACCATGGAGTGGCGCTGCTCGTGG 226
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 440 leIleIleHisGluValIleGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaG 460
 QY 227 GAGACCTGAAGCTGGTCAATCAAT----- 249
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 460 lYAspGlnIleGluGluValAsnGlyIleAspLeuArgLysAlaThrHisAspGluAlaI 480
 QY 250 -----GAACCCAGCGCTGCTCTGCTCTGTTGTGATGCCATTCGGCGGTGA 292
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 480 leAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyArgAspGluA 499
 QY 293 TCCCACTGAACGACGAGTGGATATGATCAGCTGACCCCGCGCTCCAGGAAGCTGA 352
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 499 laProTyTLySLuGluGluValCysAspThrLeuThr----- 511
 QY 353 AGGAGGTGGCTGCTGACCGCTGTCACCCGAGCGCTCGGCTGAGTGTGGTGGTGGCC 412
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 512 ----IleGluLeuGlnLysLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA 530
 QY 413 TGGAGTTGGCTGGGGCTCTTCATCTCCACCTCATCAAGGCGGTGAGCAGCAGCG 472
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 530 rGAsn--AspThrGlyValPheValSerAspIleValLysGlyGlyIleAlaAspAlaA 549
 QY 473 TCGGG---CTCCAGTAGGGGACGAGATCGCCGATCAATGATATTCATCTCCCTCT 529
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 549 spGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGlyGluAspValArgAsnA 569
 QY 530 GTACCCATGAGGAGGTATCAACCTCATTCGA---ACCAAGAAAACCTGTGTCCATCAAG 586
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 569 laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGlyThrValThrLeuGluV 589
 QY 587 TGAGACATCGGCTGATCCCGCTGAAAGACTCT-----CCTGATGAG----- 630

Db 589 alGlyArgIleIysAlaGlyProPheHisSerGluArgProSerGlnSerSerGlnv 609
 QY 631 -----
 Db 609 alSerGluGlySerLeuSerSerPheThrPheProLeuSerGlySerSerThrSerGluS 629
 QY 659 TTGTGTCGGAATCTGGGGCGTCGAGCGACCTGGGCTCC-----CCTGGAATCGGG 712
 Db 629 erLeu--GluSerSerLysLysAsnAlaLeuAlaSerGluIleGlnGlyLeuArgT 648
 QY 713 AAAACAAGGAGAAGAGCTTTCATCAGCCTGGTAGGCTCCCGAGGCTTGGCTGCAGCA 772
 Db 648 hrValGluMetLysLys-----GlyProThrAspSerLeuGlyIleSerI 663
 QY 773 TTTCAGCGGC-----CCATCCAGAAGCCTGGCATCTTTATCAGCCATGTGAAC 823
 Db 663 leAlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetMetHisP 683
 QY 824 CTGGCTCCCTCTGCTGCTGAGTG---GGATTGGAGATAGGGGACGAGATTCTCGAAGTCA 880
 Db 683 roThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleC 703
 QY 881 ATGGCTCGACTCTCTAACCTGATCACAGAGGCTGTAAATGCTGCTGAAAAATAGCC 940
 Db 703 ysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValAsnLeuLeuLysAsnAla- 722
 QY 941 GCAGCCTGACCATCTCCATTGTAGCTGAGCTGGCGGGAGCTGTTTCATGACAGACCGGG 1000
 Db 723 --SerGlySerIleGluMetGlnValAlaGly-GlyAspValSerValValThrGly 741
 QY 1001 AGCGCTGGCAGAGCGGGCAGCTGAGCTGCAGCGCAGGAGCTTCTCATGCAGAGC 1060
 Db 741 ----- 741
 QY 1061 GGCTGGGATGGATCCCAAGAAGATCTCCAGAGCAGCAGAGATGGAGCGCAAGGA 1120
 Db 741 ----- 741
 QY 1121 GAAAGAAATTGCCAGAGCAGCAGAGAGAAATGAGAGATCCGGAGGAGATGGAAAC 1180
 Db 741 ----- 741
 QY 1181 AGATTGTAGAGGAGAGAGAAGATTAAAGAACAATGGAGAAGACTGGGGCTCAAAGG 1240
 Db 741 ----- 741
 QY 1241 AACAGCTACTCTGCTTAACCATCATCTGCTGAGGTACACCCAGTACCCCTTCGCAAGC 1300
 Db 742 -----HisGlnGlnGluProAlaSerSer 749
 QY 1301 CAAAGTGATTCCGGAATATGAGNAGGCTTGACCCCTACTCTATGTTCCACC----- 1354
 Db 750 SerLeuSerPhe-----ThrGlyLeuThrSerSerSerIlePheGlnAspAsp 765
 QY 1355 -----CCAGAGCAGATCATGGGAAGGATGTCGGCTCCCTACGATCAAGAGGAGGGA 1408
 Db 766 LeuGlyProProGlnCys-----LysSerIleThrLeuGluArg-----GlyProAsp 781
 QY 1409 TCCTTAGACCTCGCCGGAAGCGGTGGACTCCCGCATTTGGG-----AAGGTGGTC 1462
 Db 782 GlyLeuGlyPheSerIleValGlyGlyTyrglySerProHisGlyAspLeuProIleTy 801
 QY 1463 GTTCTGTGTATGAGCGGGAGCTGCTGAGCGGCGATGGTGGCATTTGTAAGGGGAC 1522
 Db 802 VallysThrValPheAlaLysGlyAlaAlaSerGluAspGlyArgLeuLysArgGlyAsp 821
 QY 1523 GAGATCATGGCAATCAACCGCAAGATTGTGACAGACTACACCCCTGGCTGGCTGAGCGCT 1582
 Db 822 GlnIleIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaValAla 841
 QY 1583 GCCCTGCAGAG 1594
 |||:|||||

Db 842 IleLeuLysArg 845

Search completed: March 21, 2003, 12:44:18
 Job time : 73.8072 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	263	6.6	2037	4	US-09-306-998-3		Sequence 3, Appli
2	250.5	6.3	1881	4	US-09-233-085-3		Sequence 3, Appli
3	229.5	5.8	724	4	US-09-562-737-21		Sequence 21, Appli
4	211	5.3	724	4	US-09-562-737-28		Sequence 28, Appli
5	210.5	5.3	724	4	US-09-562-737-23		Sequence 23, Appli
6	206.5	5.2	1050	3	US-09-045-633-50		Sequence 50, Appli
7	204.5	5.1	2466	3	US-09-080-855-12		Sequence 12, Appli
8	204.5	5.1	2466	5	PCF-US94-09943-2		Sequence 2, Appli
9	204.5	5.1	2485	4	US-09-290-640-46		Sequence 46, Appli
10	203.5	5.1	1050	3	US-09-045-632-49		Sequence 49, Appli
11	202	5.1	1112	3	US-09-045-632-2		Sequence 2, Appli
12	202	5.1	1112	3	US-09-045-633-3		Sequence 3, Appli

Db	1572	SerArgSerSerThrProAlaIlePheAla-Ser---AspProAlaThrCysProIleI	1590
Qy	122	TCCGGCATAAGGTGGATTCTCTGATTGAAATGATGC	158
Db	1590	eProGly-----CysGluThrThrIleGluIleSerIy	1601
Qy	159	----AGAGAGGACATCTCTATGATGTCGTG	186
Db	1601	sGlyArgThrGlyLeu-SerIleValGlyGlySerAspThrLeuLeuGlyAlaI	1621
Qy	187	-----CGAATGTACCACAG-----ACCATGGAGCTGGCGCTCTCGTGG	226
Db	1621	IleIleHisGluValTyrGluGluGlyAlaAlaCysAspGlyArgLeuThrPalag	1641
Qy	227	GAGACCTGAAGTGTGCATCAAT-----	249
Db	1641	lyAspGlnIleLeuGluValAsnGlyIleAspLeuArgLysAlaThrHisAspGluAlaI	1661
Qy	250	-----GAAACCCAGCGCTCGCTCTCTTTTGATGCGCATTCGGCGCGCTGA	292
Db	1661	IeAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyArgAspGluA	1680
Qy	293	TCCCACTGAGCACACAGGTGGATATGATCATCAGCTACACCCCGCGCTCCAGGAAGCTGA	352
Db	1680	laProTyLysGluGluGluValCysAspThrLeuThr-----	1692
Qy	353	AGGAGTGCCTCTGGACCGCTGCACCCCGAAGGCGCTCGCCTGAGTGTGCGTGGTGGCC	412
Db	1693	----IleGluLeuGlnLysLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA	1711
Qy	413	TGGATTTGGCTCTGGGCTTCTATCTCCCACTCATCAAGGCGGTACAGCAGACAGCG	472
Db	1711	rgAsn---AspThrGlyValPheValSerAspIleValLysGlyIleAlaAspAlaA	1730
Qy	473	TCGGG---CTCCAGGTAGGGACGATCGTCCGATCAATGGATATTCATCTCCTCT	529
Db	1730	spGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGlyGluAspValArgAsnA	1750
Qy	530	GTACCCATGAGGAGGTCACTCAACTCATTCGA-----ACCAAGAAACATGTCTCCATCAAG	586
Db	1750	laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGlyThrValThrLeuGluV	1770
Qy	587	TGAGACATCGGCCTGATCCCGTGAAAGCTCT-----CTGTGATGAG-----	630
Db	1770	alGlyArgIleLysAlaGlyProPheHisSerGluArgProSerGlnSerSerGlnV	1790
Qy	631	-----CCCTCCACTGGCAGTATGTGGTCACT	658
Db	1790	alSerGluGlySerLeuSerSerPheThrPheProLeuSerGlySerSerThrSerGlnS	1810
Qy	659	TTGTGTCGGATCTGGGGCGTCGAGGCAGCTGGGTCC-----CCTGAAATCGGG	712
Db	1810	erLeu---GluSerSerSerLysLysAsnAlaLeuAlaSerGluIleGlnGlyLeuArgT	1829
Qy	713	AAAAAAGGAGAAAGGCTTTCATCAGCCTGGTAGCTCCCGAGGCGCTGGCTGCAGCA	772
Db	1829	hrValGluMetLysLys-----GlyProThrAspSerLeuGlyIleSerI	1844
Qy	773	TTTCCAGGGG-----CCATCCAGACGCTGGCATCTTTATCAGCATCTGGAAC	823
Db	1844	IeAlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetMetHisP	1864
Qy	824	CTGGCTCCCTCTCTCTGAGGTG-----GGATTTGGAGATAGGGACACAGATTCCTCAAGTCA	880
Db	1864	roThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleC	1884
Qy	881	ATGCGCTCGACTTCTCTAACTCGATCACAAGGAGGCTGTAATGTGCTGAAAAATAGCC	940
Db	1884	ysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValAsnLeuLeuLysAsnAla-	1903
Qy	941	GCAGCCTGACCATCTCCATTTGATGTAGCTGCAGCTGGCGGAGGCTGTCTATGATCAGACGGG	1000


```

QY 1636 CGCTGCCCCCAAGGAGTATGACGATGAGCTGACCTCTTCTGCTGAAGTCCAAAGGGG 1695
Db 592 -----AspIleGlnAlaHisLysPheIleGluAlaGlyGlnTy 604
QY 1696 AAACCAAAATTCACGGTGTAGGAACAGT-----GAGCTCGG 1731
Db 604 rAnSer---HisLeuTyrGlyThrSerValGlnSerValArgGluValAlaGluGlnG1 623
QY 1732 GCGCCACCTCGTGAACAAAGCCCTCGGACCAGCTTGAGAGAGGCCACATGACACACAC 1791
Db 623 yIyShiScysIleLeuAspValSerAlaAsnAlaValArgGluGlnAlaAlaHisLe 643
QY 1792 CAGATGGCATCTTGGGACCTGAATCTATCACCAGGAATCTCAAACTCCCTTTGGCCCT 1851
Db 643 u-----HisProIleAlaIlePheIleArgProArgSerLeuGluAsnValLeuGluI1 661
QY 1852 GAAC-----CAGGGCCAGATAAGGAACAGCTCGGGCCACTTTTTTGA 1893
Db 661 eAsnLysArgIleThrGluGluGln-AlaArgLysAlaPheAspArgAlaThrLysLeuG 681
QY 1894 AG-----GCCAATGTGGAGGAAGGAGGAGCAGCCAGCCGTTT 1929
Db 681 LuGlnGluPheThrGluCysPheSerAlaIleValGluGlyAspSerPheGluGluIleT 701
QY 1930 GCGAGAGATCTCAAG---ATCCAGACTCTCATTT---CCTTCTCTCTGG 1973
Db 701 yHisLysValLysArgValIleGluAspLeuSerGlyProTyrIleTrp 717

```

RESULT 4

```

US-09-562-737-28
; Sequence 28, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: US/0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-28

```

```

Alignment Scores:
Pred. No.: 5,91e-10 Length: 724
Score: 211.00 Matches: 139
Percent Similarity: 39.74% Conservative: 101
Best Local Similarity: 23.01% Mismatches: 205
Query Match: 5.29% Indels: 159
Db: 4 Gaps: 35

```

US-09-502-945-5 (1-2162) x US-09-562-737-28 (1-724)

```

QY 250 GAACCCAGCGCTGCTGCTCTGTTGATGCCATTCGG-----CCGCTGATC 294
Db 107 GlnGlnGlyArgLeuArgValAsnAspSerIleArgPheValAsnGluValAspValArg 126
QY 295 CCACTGAACACACAGGTGGAATATGATCAGCTGACCCCG----- 333
Db 127 GluSerThrHisSerAlaAlaValGluAlaLeuThrGluAlaGlySerIleValArgLeu 146
QY 334 -----CGGCGC-----TCCAGGAAGCTGAAGAGGTGCGTCTGGACCGCTCG 375
Db 147 TyrTrpMetArgArgLysProProAlaGluLysTyrIleGluLysLeuIleLys--- 165

```

```

QY 376 CACCCCGAAGCCCTCGGCTGAGTGTGGTGGTGGC-----CTGAGATTT 420
Db 166 GlyProAlaGlyLeuGlyPheSerIleAlaGlyAspGlyAsnGlnHisIleProGly 185
QY 421 GGCTGTGGCTCTCATCTCCCATCATCAAGCGCTCAGGCAGACAGCGTGGG--- 477
Db 186 AspAsnGluIleTyrValThrLysIleIleGluGlyPheAlaAlaHisLysAspGlyArg 205
QY 478 CTCAGGTGAGGGAGAGATCGTCCGGATCAATGATATTCATCTCTCTCTGACCCAT 537
Db 206 LeuGlnGlyGlyAspLysIleLeuAlaValAsnSerHisGlyLeuGluAspValMetHis 225
QY 538 GAGGAGTCAATCAACCTCATTCGA---ACCAGAAGAACTGTGTCCATCAAAAGTGAAGAC 594
Db 226 GluAspIleValAlaAlaLeuLysAsnThrTyrAspLysValTyrLeuLysValAlaLys 245
QY 595 ATCGGCTCTCCCGTGAAGAGCTCT-----CCTGATGAGCCCTCACT---TGG 642
Db 246 ProSerLeuAlaTyrLeuSerAspSerTyrAlaProMetAspIleThrThrSerTyrSer 265
QY 643 CAGTATGTGGATCATGTTGTCTCGGAATCTGGGGCGGTGCGAGGACG----- 690
Db 266 GlnHisAsnAspAsnGluIleSerHisSer---SerTyrGlnGlyThrAspTyrProThr 284
QY 691 -----CTGGCTCCCTCGAAATCGGAA----- 714
Db 285 AlaMetThrArgThrSerProArgArgTyrSerProValSerLysAspLeuGlyGlu 304
QY 715 -----AACAAGGAAGAAGTCTTCATCAGCTGTGAGGTCTCCGAGTCTGGC 765
Db 305 GluAspIleThrArgGluProArgArgIleValIleHisValGlySerThrGlyLeuGly 324
QY 766 TGCAGCATTTCCAGGGGCCCATCCAGAACCTGGCATCTTTTATCAGCCATGTG----- 819
Db 325 PheAsnIleTrpGlyGlyGlu---AspGlyGlyGlyIlePheTyrSerPheIleLeuAla 343
QY 820 ---RAACCTGCTCCCTGCTGCTGAGTGGATTGGAGATAGGGACACCATGTGCGAA 876
Db 344 GlyGlyProAlaAlaLeuSerGlyGlu-----LeuArgLysGlyAspAspIleLeuSer 361
QY 877 GTCATATGCGCTCGACTTCTTAACCTGGATCACAAGGAGGCTGTAAATGTCTGAAAAAT 936
Db 362 ValAsnGlyValAspLeuGluAsnAlaSerHisGluGlnAlaAlaIlePheLeuLysAsn 381
QY 937 AGCCGACCTTGACCATCTCCATTTGATCTGACGTGCGGCGGAGCTGTTTCATGACAC 996
Db 382 Ala---GlyGlnThrValThrGlyIleAlaGln-----TyrLysProGlu 395
QY 997 CGGAGCGGCTGGCAGAGCGCGCAGCGTGTGAGTGTGCGGCGGAGGAGCTTCTCATGCAG 1056
Db 396 GluTyrHisArgPheGluAlaLysIleHisAspLeuIleGluGln-----LeuMetAsn 413
QY 1057 AAGCGGCTG-----GCGATGGAGTCCAAACAAGATCCTCCAG----- 1092
Db 414 SerSerLeuGlyLysGlyThrAlaSerLeuArgSerAsnProLeuArgGlyPheTyrIle 433
QY 1093 -----GAGCAGCAGAGATGGAG 1110
Db 434 ArgAlaLeuPheMetTyrAspLysThrLysAspCysGlyPheAsnSerGlnAlaLeuSer 453
QY 1111 CGGCAAGAGAGAAAGAAATTCGCCAGAGGACAGGAGGAAATGAGAGATACCGG--- 1167
Db 454 PheHisPheGlyGlnValLeuHisValIleAspAlaSerAspArgGluTrpIleAla 473
QY 1168 -----AAGGAGATGGAACAGATT-----GTAGAGGAGAGAGAGAG 1203
Db 474 ArgArgValHisThrAspSerGluThrAspAspIleGlyPheValProSerLysArg 493
QY 1204 TTTAAGACCAATGG-----GAAGAAGACTGG----- 1230
Db 494 ValGluArgArgTrpSerArgLeuLysAlaLysAspTrpTyrSerSerSerGlySer 513
QY 1231 ---GGCTCAAGGAACAGCTACTCTTGGCTTAAACCATCACT---GCTGAGGTACAC--- 1281

```



```
QY 1237 -----AAGGAACAGCTACTTGTGCTAAACCATCACT---GCTGAGTA 1278
Db 512 GlyGluGlnGlyArgGluAspSerValLeuSerPheGluThrValThrGlnMetGluVal 531
QY 1279 CACCCAGTACCCCTTCGCAAGCCAAAGTGATTTCCGGAATATGAGGAAGGCTTTGACCC 1338
Db 532 HisGlyAlaArg-----Pro 536
QY 1339 CTACTCTATGTTACCCAGGAGGAGATCATGCGGGAAGGATGTCGGGTCCTACGATCAA 1398
Db 537 IleIleLeuGlyProHisLysAspArgAla-----547
QY 1399 GAAGGAGGATCCTTAGACCTGGCCCTGGAAGGGGTGTGGACTCCGCCATTTGGGAAGGT 1458
Db 548 AsnAspAspLeuLeuIleGluPheProAspLys-----558
QY 1459 GGTGCTTTCTGCTGTATGACGGGGAGCTGCTGAGCGGATGTGGCATTGTGAAGG 1518
Db 559 -----PheGlySerCys-LysProHisThrThrArgProLysAr 571
QY 1519 GGACGAGATCGCAATCAACGCAAGATTGTGACAGACTACACCTGGCTGAGGCTGA 1578
Db 571 gGlu-----LeuGluIleAspGlyArg-----AspTyrHisPheMetSerSerAr 586
QY 1579 CGTGCCTCGAGAGGCTGGAATCAGGCGGGGACTGGATCGACCTTGTGGTTGCCGT 1638
Db 586 gGluLysMetGluLys-----591
QY 1639 CTGCCCCCAAGGAGTATGACGATGAGCTGACCTTCTGCTGAAGTCCAAAGGGGAAA 1698
Db 592 -----AspAsnGlnAlaHisLysPheIleGluAlaGlyArgTyrAs 605
QY 1699 CCAATTCACCGCTTAGAACAAGT-----GAGCTCCGGGCC 1734
Db 605 nSer---HisLeuTyrGlyThrSerSerGlnSerValArgGluValAlaGluGlnThrLy 624
QY 1735 CCACCTCGTGAACAAAGCCTCGGACCGCTTGAGAGCGCCACATGACACACACAGG 1794
Db 624 SHiscysIleLeuAspValSerAlaValAlaValArgLeuGlnAlaHis-----642
QY 1795 ATGCATCCTTTGGACCTGAATATCACCCAGGAATCTCAAACTCCCTTTGGCCCTGAA 1854
Db 643 -TriPheProIleAlaIlePheIleArgProTyr-SerLeuGluAsnValLeuGluIleAs 662
QY 1855 C-----CAGGGCCAGATAAGGAACAGCTCGGGCCACTTTTGAAG- 1895
Db 662 nAlaArgIleThrGluGluGln-AlaArgLysAspPheAspArgAlaThrLysLeuGluG 682
QY 1896 -----GCCAATGTGGAGGAAGGAGCAGCCAGCCCTTTGGG 1932
Db 682 InPhePheThrGlucysPheSerAlaIleValGlyLysSerPheGluGluIleTyrH 702
QY 1933 AGAAGATCTCAAGATC-----CAGACTCTCATCTCTCTCTCTCTG 1973
Db 702 IsHisValLysArgValIleGluAspLeuSerIleProTyrIleTyr 717
```

RESULT 6
US-09-045-632-50
; Sequence 50, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Haganir, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA

ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/045,632
FILING DATE: 19-MAR-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/041,016
FILING DATE: 19-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Corless, Peter F.
REGISTRATION NUMBER: 33,860
REFERENCE/DOCKET NUMBER: 48147/1699-CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1050 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-045-632-50

Alignment Scores:
Pred. No.: 1,81e-09 Length: 1050
Score: 206.50 Matches: 189
Percent Similarity: 35.22% Conservative: 99
Best Local Similarity: 22.21% Mismatches: 229
Query Match: 5.18% Indels: 229
DB: 3 Gaps: 35
US-09-502-945-5 (1-2162) x US-09-045-632-50 (1-1050)
QY 322 CAGCTGACCCCGCGCTCCAGGAAGCTGAAGAGGTGCTGTCGACCGCTGCACCCC 381
Db 77 GlnSerIleProGluGluPheArgGlyIleThrMetValGluLeuIleLysArgGluGly 96
QY 382 GAAGCCCTCGGCTGAGTGTGCTGGTGGCTGGAGTTGGCTGGCTTCATCTCC 441
Db 97 SerThrLeuGlyLeuThrIleSerGlyGlyThrAspLysAspGlyLysProArgValSer 116
QY 442 CACCTCATCAAGGCGGT---CAGGCAGACAGCTCGGGCTCCAGGTAGGGGACGAGATC 498
Db 117 AsnLeuArgProGlyGlyLeuAlaAlaArgSerAspLeuLeuAsnValGlyAspTyrIle 136
QY 499 GTCCGGATCAATGATATTCATCTCTCTCTACCCATGAGGAGGTCAACACCTCAT 558
Db 137 ArgSerValAsnGlyIleArgLeuThrArgLeuArgHisAspGluIleIleThrLeuLeu 156
QY 559 CGAAC-----564
Db 157 LysAsnValGlyGluArgValValLeuGluValGluTyrGluLeuProProAlaPro 176
QY 565 -----AAGAAACGTGTCTCCATCAAGTG-----AGA 591
Db 177 GluAsnAsnProArgIleIleSerLysThrValAspValSerLeuTyrLysGluGlyAsn 196
QY 592 CACATCGGCTGATCCCGCTGAAAGCTCTCCTGATGAG-----CCCCC 636
Db 197 SerPheGlyPheValLeuArgGlyGlyAlaHisGluAspLeuHisLysSerArgProLeu 216
QY 637 ACTGGCAGTATGTG-----651
Db 217 ValLeuThrTyrValArgProGlyGlyProAlaAsnArgGluGlySerLeuLysValGly 236
QY 652 GATCAGTTTGTGCGGAATCTGG-----675
||||| :|||

Db 237 AspArgLeuLeuSerIle***GlyIleProLeuHisGlyAlaSerHisAlaThrAlaIle 256
QY 676 -----GCGTGGCGAGGCGAGCTGGGCTCC 699
Db 257 AlaThrLeuGlnGlnCysSerHisGluAlaLeuPheGlnValGluTyrAspValAlaThr 276
QY 700 CCTGAAATCGGAAACAAGAGGAGGAGTCTTCATCAGCCCTGGTAGGCTCCGAGGC 759
Db 277 ProAspThrValAlaAsnAlaSerGlyProLeuValGluIleAlaLysThrProGly 296
QY 760 -----CCTGGCTGCAGATTCCTCCAGCGGCCCATCCAG---AAGCCTGGCATCTTTATC 810
Db 297 SerAlaLeuGlyIleSerLeuThrThrGlySerHisArgAsnLysProAlaIleThrIle 316
QY 811 AGCCATGTGAACCGGCTCTCTGCTGCTGAGTGGGA---TTGGAGATAGGAGACCA 867
Db 317 AspArgIleLysProAlaSerValAlaAspLysAsnGlyAlaLeuHisAlaGlyGluHis 336
QY 868 ATTGTGGAAGTCAATGGCGTGCAGTCTCTTAACCTGGATCAAGAGGAGGCTGTAATGTG 927
Db 337 IleLeuAlaIleAspGlyThrSerThrGluHisCysSerLeuValGluAlaThrLysLeu 356
QY 928 CTGAAAT---AGCGGAGCGTACCATCTCCATTGTAGTGCA-----GCTGGCGG 978
Db 357 LeuAlaSerValThrGluLysValArgLeuGluIleLeuProAlaProGlnSerArgArg 376
QY 979 GAGCTGTTTCATCAGACAGCGGCGGCTGCGAGAGCGCGCGCGTGC--- 1027
Db 377 ProLeuLysProGluAlaValArgIle-GlnArgSerGluGlnLeuHisHisTrpAs 396
QY 1028 -----AGCTGC---AGCGGAGGAGCTCTCATCAGCAAGCGGCTGGCGATG 1071
Db 396 pProCysValProSerCysHisSerProArgProSerHisCysArgAlaPro---ThrTr 415
QY 1072 GAGTCCA-----ACAAGATCTCCAGGAGCAGCAGGATGGAGCGC----- 1114
Db 415 pAlaProGlyGlyGlnAspGlnSerArgSerValSerThrProPheSerSerProTh 435
QY 1115 -----AAAGGAGAAAGAAATTCGCCAGAGAGGCGAGCAGAGGAA 1152
Db 435 rMetAsnProAlaPheProCysAlaAsnAlaSerThrLeuProArgGlyProMetSerPr 455
QY 1153 AATGAGAGATACCGAAGAGATGGAACAGATGTAGAGGAGGAGAGAGA----- 1201
Db 455 oArgThrAlaGlyArgArgGln-----ArgArgLysGluHisArgSerSe 472
QY 1202 ----AGTTTAAGAACAAATGGGAAGAGACTGGGGCTCAAGAGAACAGACTACTCTGCT 1257
Db 472 rLeuSerLeuAlaSerSerThrValGlyProGlyGlyGln----- 485
QY 1258 AAACCATCCTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTGATTCGCGAA 1317
Db 486 -----IleValHisThrGlu----- 490
QY 1318 ATATGAGGAGGCTTTGACCCCTACTCTATGTTCACCCAGAGCAGATCATCGGGAAGA 1377
Db 491 -----ThrThrGluValValLeuLeuCysGlyAs 499
QY 1378 TGTCGGCTCTACGCATCAAGAAGGAGGATCTTTAGACCTGGCCCTGGAAGCGGGTGT 1437
Db 499 p-----ProLeuSerGlyPheGlyLeuGlnLeuGlnGlnGlyIle 512
QY 1438 G-----GACTCCCCCATTTGGGAAGGTGGTTCGTCTGTGTATGA 1479
Db 512 ePheAlaThrGluThrLeuSerSerPro-----ProLeuValArgPheIleGluPr 529
QY 1480 CGGGGAGCTGTGACCGGATGTGGCATTTGGAAGGGGAGGATCATGGCAATCAA 1539
Db 529 oAspSerProAlaGluArgCysGlyLeuLeuGlnValGlyAspArgValLeuAlaIleAs 549
QY 1540 CGCAAGATGTGACAGACTACACCTGCTGAGGCT-----GACGCTGC 1584
Db 549 nGlyIleAlaThrGluAspGlyThrMetGluGluAlaAsnGlnLeuLeuArgAspAlaI 569

QY 1585 CCTCAGAAAGCGCTGAATCAGGCGGGGACTGATGCACCTTGTGGTCCCTGCTGCC 1644
Db 569 aLeuAlaArgLysIleValLeuGluIleGluPhe---AspValAlaSerValIlePr 588
QY 1645 C-----CCAAAGGAGTATGACGATGACGATGACCTTCTT 1677
Db 588 oSerSerGlyThrPheHisValLysLeuProLysArgGlyValGluLeuGlyIleTh 608
QY 1678 GCTCAAGTCC-----AAAAGGGAAACCAATTCACGCGTGTAGGAAACAGTGA 1728
Db 608 rIleSerSerAlaSerArgLysArgGly----- 617
QY 1729 CCGGCCCCACCTCGTGAACACAAAGCCCTCGGACCAAGCTTTCAGAGAGCCACATGACACA 1788
Db 618 -GluProLeuIleIle-----SerAspIleLysLysGlySerValAlaH1 632
QY 1789 CACC-----AGATGGCATCTTGGGACCTGATCTATCACCAGGAATCTCAACTC-- 1840
Db 632 sArgThrGlyThrLeuGluProGlyAspLysLeuLeuAlaIleAspAsnIleArgLeuAs 652
QY 1841 -----CCTTTGGCC-----CTGAACACAGGG 1860
Db 652 pHisCysProMetGluTyrAlaValGlnIleLeuProGlnCysGluAspLeuValLysLe 672
QY 1861 CCAGATAAGGAACAGCTCGGCGCACCTTTTGAAGGCCAATGTGGAGA----- 1909
Db 672 uLysIleArgLysAspGluAspAsnSerAspGluGlnGluSerSerGlyAlaValSerTy 692
QY 1910 -----AAGGAGCAGCCAGCCGTTTGGGA-----GAAGA 1938
Db 692 rThrValGluLeuLysArgTyrGlyProLeuGlyIleThrIleSerGlyThrGluG1 712
QY 1939 TCTCAAGGATCAGACACTCATCTCTTCTTCTGCGCCAGTGAATTTGGTCTCTCCAGC 1998
Db 712 uProPheAspPro-IleIleIleSerGlyLeu-----ThrLysArgGlyLeuAlaGluA 730
QY 1999 TTGCGGAGCTCTCTCTTGAACCTTAATAGACCCAGCTGAGTCTCTCTCTCCATC 2058
Db 730 rgThrGlyAlaSerMetLeuGlyThrAlaTyrTrpProSer***SerValSerLeuLysG 750
QY 2059 CCTCTCTCTCCCTCTCTCTAATTTGCTGCGCAGGATTTGCTCAACCTTACTCTG 2117
Db 750 lYArgProLeuSerGluAlaIleHisLeuLeuGlnValAlaGlyGluThrValThrLeu 769

RESULT 7

; Sequence 12, Application US/09080855A
; Patent No. 6083721
; GENERAL INFORMATION:
; APPLICANT: Saras, Jan
; APPLICANT: Franz, Petra
; APPLICANT: Aspenstrm, Pontus
; APPLICANT: Hellman, Ulf
; APPLICANT: Gopez, Leonel Jorge
; APPLICANT: Heldin, Carl-Henrik
; TITLE OF INVENTION: PARG, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPL1
; FILE REFERENCE: L0461/7030
; CURRENT APPLICATION NUMBER: US/09/080,855A
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: 08/805,583
; EARLIER FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2466
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-080-855-12

Alignment Scores: 4.43e-09 Length: 2466
Pred. No.: 204.50 Matches: 143
Score:

Percent Similarity: 36.35% Conservative: 94
 Best Local Similarity: 21.93% Mismatches: 214
 Query Match: 5.13% Indels: 29
 DB: 3 Gaps: 201

US-09-502-945-5 (1-2162) x US-09-080-855-12 (1-2466)

```

QY 1 CTGGCCGCTGCGGCTCGCGCTCTT----- 27
DB 991 ProProProGlnThrValAlaGluLeuValGlyLysProSerHisGlnMetSerArgSer 1010
QY 28 ---TCCAGTCTCTGGCAGCGGCGCCACCGAAGGAGCGGCTCGTGAACGACGACGCTGGA 84
DB 1011 AspaIaGluSerLeuAlaGlyValThrLysLeuAsnAsnSerLysSerValAla----- 1028
QY 85 CTGGCCGACGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 144
DB 1029 -----SerLeuAsnArgSerProGluArgArg---LysHisGluSerAspSerSer 1044
QY 145 ATTGAAATGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 204
DB 1045 SerIleGluAspProGlyGlnAlaTyrValLeuAspValLeu-----HisLysArg 1061
QY 205 ATGGAGCTGGCCGCTGCTGCTGGGAGACCTGAAGCTGGTCAATGAACCCAGCGCTCTG 264
DB 1062 TrpSer-----IleValSerSerProGluArg--- 1070
QY 265 CTTCTGTTGATGCCATNTCCGCGGTGATCCACGTAAGCAGCAGGCGGAATATGATCAG 324
DB 1071 -----GluIleThrLeuValAsnLeuLysLysAspAlaLysTyr----- 1083
QY 325 CTGACCCCGCGCTCCAGGAGCTGAAGAGGCTGCTGACCGCTGACACCCGAA 384
DB 1083 ----- 1083
QY 385 GGCCTCGGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 435
DB 1084 GlyLeuGlyPheGlnIleIleGlyGlyGlyLysMetGlyArgLeuAspLeuGlyIlePhe 1103
QY 436 ATCTCCACCTCATCAAGCGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 492
DB 1104 IleSerSerValAlaProGlyGlyProAlaAspPheHisGlyCysLeuLysProGlyAsp 1123
QY 493 GAGATCGTCCGATCAATGGATATTCATCTCTCTGATACCATCAGGAGGATCATCAAC 552
DB 1124 ArgLeuIleSerValAsnSerValSerLeuGluGlyValSerHisAlaAlaIleGlu 1143
QY 553 CTCATTGCAACCAAGAAACTGTGTCCATCAAGATGAGACACATCGCGCTGATCCCGTG 612
DB 1144 IleLeu----- 1145
QY 613 AAAAGCTCTCTGATGAGCCCTCACTTGGCAGTATGTGATGATGATGATGATGATGATGAT 672
DB 1146 GlnAsnAlaProGluAsp-----ValThrLeuValIleSerGlnPro 1159
QY 673 GGGGCGTCCGAGGAGCGCTGGCTCCCTGGAATCGGAAACAAAGAAAGAGGTC 732
DB 1160 -----LysGluLys----- 1162
QY 733 TTTCATCAGCTGTTAGGCTCCGAGCGCTTGCTGCAGCATTTCCAGCGCGCCCATCCAG 792
DB 1163 ---IleSer-----LysValProSerThrProValHis 1172
QY 793 AAGCCTGGCATCTTTATCAGCCATGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 852
DB 1173 LeuThrAsnGluMetLysAsnTyrMetLysLysSer----- 1185
QY 853 GAGATAGGGGACGATGTCGAATCAATGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 912
DB 1186 -----TyrMetGlnAspSerAlaIleAspSerSerLysAspHis 1200
QY 913 GAGGCTGTAATGTCTGAAATAAGCCGACGCTGACCATCTCCATTTGCTAGCTGACCT 972
DB 913 ----- 972

```

RESULT 8

PCT-US94-09943-2
 ; Sequence 2, Application PC/TUS9409943
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; APPLICANT:
 ; APPLICANT:

```

DB 1201 TrpSerArgGlyThrLeu-----ArgHisIleSerGluAsnSerPheGlyProSer 1217
QY 973 GGC-----CGGAGCTGTTTCATGACAGACGCGGAGCGCTGGCAGAGCGCGGACGCT 1026
DB 1218 GlyGlyLeuArgGluGlySerLeuSerGlnAspSerArgThrGluSer-AlaSerLe 1237
QY 1027 GAGCTGCAGC-----GGCAGGAGCTTCTCATGCAAGACGCGC-----TGGCGATG 1071
DB 1237 userGlnSerGlnValAsnGlyPhePheAlaSerHisLeuGlyAspGlnThrTrpGlnG 1257
QY 1072 GAGTCCAAACAGATCCTCCAGGAGCAGCAGGAGATGGAGCGCAAGAGGAGAAAGAAAT 1131
DB 1257 user-GlnHisGlySerProSerProSerValIleSerLysAlaThrGluLysGluThrP 1277
QY 1132 GCCCAGGAGCAGCAGAGGAAATGAGATACCGGAA---GGAGATGGAACAGATGTT- 1187
DB 1277 heThrAspSerAsnGlnSerLysThrLysLysProGlyIleSerAspValThrAspTyrS 1297
QY 1188 -----AGAGGAGGAGAGAGAGTAAAGAGCAATGGGAAGAGACTGGGGCTCAAGGAA 1242
DB 1297 erAspArgGlyAspSerAspMetAspGluAla----- 1308
QY 1243 CAGCTACTCTTGCCTAAACCATCCTGCTGAGGTACACCCAGTACCCCTTCGCAAGCA 1302
DB 1308 hrTyrSerSerSerGlnAspHisGln-----ThrProLysGlnLysSerSerS 1325
QY 1303 AAGTGATTTCCGGAATATGAGGAAGGCTTTGACCCCTTACTCTATGTTACCCCGCAGCA 1362
DB 1325 erValAsnThrSerAsn--LysMetAsnPheLysThrPheSerSerProLysPr 1344
QY 1363 GATCATGGGAGGATGTCGCGCTCTACGATCAAGAGGAGGAGGATCTTAGACCTGGC 1422
DB 1344 o-----GlyAspIlePheGluValGluLeuAlaLysAsnSerLeuGlyIleSe 1362
QY 1423 CTGGAAGCGGTGGGACTCCCCCAT-----GGGAAGCTGGTCTGTTCTGCTGTGTA 1476
DB 1362 rValThrGlyGlyValAsnThrSerValArgHisGlyGlyIleTyrValLysAlaVal 1382
QY 1477 TGAGCGGAGCTGCTGAGCGGCATGTGGCATTTGTGAAGGCGGAGCAGATCATGGCAAT 1536
DB 1382 eProGlnGlyAlaAlaGluSerAspGlyArgIleHisLysGlyAspArgValLeuAlaVa 1402
QY 1537 CAACGGCAAGATTTGACAGACTACACCTGCTGAGGCTGACGCTGCCCTG----- 1588
DB 1402 lAsnGlyValSerLeuGluGlyAlaThrHisLysGlnAlaValGluThrLeuArgAsnTh 1422
QY 1588 ----- 1588
DB 1422 rGlyGlnValValHisLeuLeuGluLysGlyGlnSerProThrSerLysGluHisVa 1442
QY 1589 -----CAGAAGCGCTGGAATCAGGCGCGGAGCTG 1617
DB 1442 lProValThrProGlnCysThrLeuSerAspGlnAsnAlaGlnGlyGlnGlyProGluLy 1462
QY 1618 GATGACCTTGTGTTGCGCTGCTGCCCGCCCAAGAGATATGAC-----GATGAGCT 1668
DB 1462 sValLysLysThrThrGlnVal-----LysAspTyrSerPheValThrGluGluAs 1479
QY 1669 GACCTC-----TTGCTGAAGTCCAAAGGGGAAACCAATTCACGCTTAGGAA 1719
DB 1479 nThrPheGluValLysLeuPheLysAsnSerSerGlyLeuGlyPheSerPheSerArgGl 1499
QY 1720 CAGTGAGCTCGCGCCACCTCGCTGAAC 1747
DB 1499 uAspAsnLeuIleProGluGlnIleAsn 1508

```


Qy	325	CTGACCCCGCGCGCTCCAGGAAGCTGAAGGAGTGGCTGGACACGCTGTGCACCCCGAA	384
Db	1083	-----	1083
Qy	385	GGCTCGGCTGAGTGTGCTGGTGGCTGGAGTTGGC-----TGTGGGCTCTTC	435
Db	1084	: : : : : :	1103
Qy	436	ATCTCCCACTCATCAAAAGCGGTGAGCAGACAGCGTCGGG----CTCCAGGTAGGAGC	492
Db	1104	: : :	1123
Qy	493	GAGATCGTCGGATCAATGGATTAATCCATCTCCTCTGATCCCATGAGGAGGTCAATCAAC	552
Db	1124	: : :	1143
Qy	553	CTCATTCGAACCAAGAAAACTGTGTCATCAAGTGAACATCGGCCTGATCCCGCTG	612
Db	1144		1145
Qy	613	AAAGCTCTCTGATGAGGCCCTCACTTGGCGAGTATGTGGATCAAGTTGTGTGCGGAATCT	672
Db	1146	: : :	1159
Qy	673	GGGCGGCTGCGAGCGCCTGGCTCCCTGGAATCGGAAAAACAAGGAGGAAGGTC	732
Db	1160		1162
Qy	733	TTTCATCAGCTGTAGGCTCCGAGCGCTTGGCTGCAGCATTTCCAGCGGCCCATCCAG	792
Db	1163		1172
Qy	793	AGCGCTGGCATCTTTATACGCAATGTGAACCTGGCTCCCTGCTGCTGAGGTGGGATTC	852
Db	1173	: : :	1185
Qy	853	GAGATAGGGACCATGTTCTGAAGTCAATGCGCTGCACTTCTTAACCTGGATCAACAG	912
Db	1186	: : :	1200
Qy	913	GAGCGCTAAATGCTGTAATAATACCGCAGCGCTGACCATCTCCATTTAGCTGCAGCT	972
Db	1201	: : :	1217
Qy	973	GGC-----CGGAGCTGTTCATGACAGCCGGAGCGGTGGCGAGAGCGCGCACCGT	1026
Db	1218	: : :	1237
Qy	1027	GAGTCGAG-----GGCAGGAGCTTCTCATGCAAGAGCGC-----TGGCGATG	1071
Db	1237	: : :	1257
Qy	1072	GAGTCCAACAAGATCTCCAGGACGAGGAGATGGAGCGGCAAGAGAGAAAGAAATT	1131
Db	1257	: : :	1277
Qy	1132	GCCCAAGGACGACGAGGAAATGAGATACCGGAA---GGAGTGGAGCAAGATTGT-	1187
Db	1277	: : :	1297
Qy	1188	-----ACAGAGGAGAGAGATTAAAGAGCAATCGGGAAGAGACTGGCGCTCAAGGAA	1242
Db	1297	: : :	1308
Qy	1243	CAGCTACTCTTGCTTAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCA	1302
Db	1308	: : :	1325
Qy	1303	AAGTGATTTCGGAAATATGAGGAAGCGCTTTGACCCCTACTCTATGTTACCCACGACGA	1362
Db	1325	: : :	1344
Qy	1363	GATCATGGGAAGGATGTCCGCTCCTACGCAATCAAGGAAGAGGAGTCCCTTAGACCTGGC	1422

Db	1344	o-----GlyAspIlePheGluValGluLeuAlaLysAsnAspAsnSerLeuGlyIleSe	1362
QY	1423	CCTGGAAGCGGTGTGGACTCCCCCAATT-----GGGAAGCTGGCTGTTCTGCTGTGA	1476
Db	1362	rValThrGlyGlyValAsnThrSerValArgHisGlyGlyIleTyValLysAlaValII	1382
QY	1477	TGAGCGGGAGCTGCTGACGGCGATGGTGGCATTTGTAAAGGGACGAGATCATGGCAAT	1536
Db	1382	eProGInGlyAlaAlaGluSerAspGlyArgIleHisLysGlyAspArgValLeuAla	1402
QY	1537	CAACGGCAAGATTGTGACAGACTACACCTGGCTGAGCTCACCTGCCCTG-----	1588
Db	1402	lAsnGlyValSerLeuGluGlyAlaThrHisLysGlnAlaValGluThrLeuArgAsnTh	1422
QY	1588	-----	1588
Db	1422	rGlyGlnValValHisLeuLeuGluLysGlyClnSerProThrSerLysGluHisVa	1442
QY	1589	-----CAGAAGCCCTGGAATCAGGGCGGGGACTG	1617
Db	1442	lProValThrProGlnCysThrLeuSerAspGlnAsnAlaGInGlyGlnGlyProGluLy	1462
QY	1618	GATCGACCTTGTGGTTCGCTGCTGCCCCCAAGGACGATATGAC-----GATGAGCT	1668
Db	1462	sValLysLysThrThrGlnVal-----LysAspTyrSerPheValThrGluGluAs	1479
QY	1669	GACCTTTC-----TTGCTGAAGTCCAAAAGGGGAAACCAATTCACGCTTAGGAAA	1719
Db	1479	nThrPheGluValLysLeuPheLysAsnSerSerGlyLeuGlyPheSerPheSerArgGI	1499
QY	1720	CAGTGAGCTCCGGCCCCCCTCGTGAAC	1747
Db	1499	uAspAsnLeuIleProGluGlnIleAsn	1508

[illegible]

Db	1384	yGlyValAsnThrSerValArgHisGlyGlyIleTyrValLysAlaValIleProGlnGlu	1404
Qy	1486	AGCTGCTGAGCGGCATGTTGGCATTGTGAAAGGGGACGAGATCATGCCAATCAACGGCAA	1545
Db	1404	YAlaAlaGluSerAspGlyArgIleHisLysGlyAspArgValLeuAlaValAsnGlyIle	1424
Qy	1546	GATTGTGACAGACTACACCTCGCTGAGCTGAGCTGACCTGCCCTG	1588
Db	1424	lSerLeuGluGlyAlaThrHisLysGlnAlaValGluThrLeuArgAsnThrGlyGlnVal	1444
Qy	1588	-----	1588
Db	1444	lValHisLeuLeuLeuLysGlyGlnSerProThrSerIysGluHisValProValTh	1464
Qy	1589	-----CAGAAGCCCTGGAAATCATGGCGGGGACGTGATCGACCT	1626
Db	1464	rProGlnCysThrLeuSerAspGlnAsnAlaGlnGlyGlnGlyProGluLysValLysLy	1484
Qy	1627	TGTGGTTGGCTGCTGCCCCCAAGGAGTATGAC-----GATGAGCTGACCTTC--	1675
Db	1484	sThrThrGlnVal-----LysAspTyrSerPheValThrGluGluAsnThrPheGlu	1501
Qy	1676	-----TTGCTCAAGCTCCCAAGGGGAAACAAATTCACGCGTTAGGAAACAGTGAAGCT	1728
Db	1501	uValLysLeuPheLysAsnSerSerGlyLeuGlyPheSerPheSerArgGluAspAsnLe	1521
Qy	1729	CGCGCCCCACCTCGTGAAC	1747
Db	1521	uileProGluGlnIleAsn	1527

RESULT, T 10

US-09-045-632-49
; Sequence 49, Application US/09045632

: PATENT NO. 6001373
 : GENERAL INFORMATION:

APPLICANT: HUGANIR, Richard L.

APPLICANT: HUYANLI, RICHARD

APPLICANT: Dong, Hualing

AFRICANI: DONG, HUALING
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND

; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP-RELATED MOLECULES
; TITLE OF INVENTION: GRIP-RELATED MOLECULES

NUMBER OF SEQUENCES: 105

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

; CITY: Boston

; STATE: MA

COUNTRY: USA

ZIP: 02109

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/045,632

FILING DATE: 19-MAR-1998

CLASSIFICATION: 435

;
PRIOR APPLICATION DATA:
;
APPLICATION NUMBER:

APPLICATION NUMBER: US 60/041,016
FILING DATE: 10-MAR-1997

FILING DATE: 19-MAR-1997
ATTORNEY/AGENT INFORMATION:

ATTORNEY/AGENT INFORMATION:
NAME: Corliss Peter F

NAME: COFFESS, PETER F.
REGISTRATION NUMBER 33

REGISTRATION NUMBER: 33,
REFERENCE/DOCKET NUMBER:

REFERENCE/DOCKET NUMBER: 4814//1055-CIF
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEPHONE: 617-523-5400
TELEFAX: 617-523-6440

; INFORMATION FOR SEQ ID NO: 4

SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1050 amino aci

```

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: pro

US-09-045-632-49

Db 377 ProLeuLysProProGluAlaValArgIle-GlnArgSerGluGlnLeuHisHisTrpAs 396
QY 1028 -----AGCTGC-----ACGGCAGGAGCTTCTCATGCAGAACGGCTGCGCATG 1071
Db 396 pProCysValProSerCysHisSerProArgProSerHisCysArgAlaPro---ThrTr 415
QY 1072 GAGTCCA-----ACAAGATCCTCCAGGAGCAGCAGAGATGAGCGGC-----1114
Db 415 pAlaProGlyGlnAspGlnSerArgSerValSerSerThrProPheSerSerProTh 435
QY 1115 -----AAAGGAAAAGAAATGTCAGAGAGCCAGCAGAGGAA 1152
Db 435 rMetAsnProAlaPheProCysAlaAsnAlaSerThrLeuProArgGlyProMetSerPr 455
QY 1153 AATGAGAGATACCGAAGGAGATGAGACAGATGTGAGGAGGAGAGA-----1201
Db 455 oArgThrThrAlaGlyArgArgGln-----ArgArgLysGluHisArgSerSe 472
QY 1202 ----AGTTTAAAGCAATGGGAAGAGACTGGGCTCAAGAGCAAGCAGCTACTCTGCCT 1257
Db 472 rLeuSerLeuAlaSerSerThrValGlyProGlyGln-----485
QY 1258 AAAACCATCACTGTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTGTATTCGCGAA 1317
Db 486 -----IleValHisThrGlu-----490
QY 1318 ATATGAGGAAGCTTACCCCTACTATGTTACCCAGCAGAGATCATGGGAGGA 1377
Db 491 -----ThrThrGluValValLeuCysGlyAs 499
QY 1378 TGTCGGGCTCTACGCATCAAGAGGAGGATCTTAGACTGCGCTGGAAGGGGTGT 1437
Db 499 p-----ProLeuSerGlyPheGlyLeuGlnLeuGlnGlyGlyI 512
QY 1438 G-----GACTCCCCCATTTGGGAAGGTGCTGCTTCTGCTGTATGA 1479
Db 512 ePheAlaThrGluThrLeuSerPro-----ProLeuValArgPheIleGluPr 529
QY 1480 GCGGGAGCTCTGAGCGCATGTGCTGATGTAAGGGGACAGATCATGGCAATCAA 1539
Db 529 oAspSerProAlaGluArgCysGlyLeuLeuGlnValGlyAspArgValLeuAlaIleAs 549
QY 1540 CGGCAAGATTGTGACAGATACACCTGGCTGAGGCT-----GACGCTGC 1584
Db 549 nGlyIleAlaThrGluAspGlyThrMetGluGlnAlaAsnGlnLeuLeuArgAspAlaAl 569
QY 1585 CCTCAGAGCCCTGGAATCAGGCGGGGACTGATCGACTGTGTTGCTGCGCTGCC 1644
Db 569 aLeuAlaArgLysIleValLeuGluIleGluPhe---AspValAlaGluSerValIlePr 588
QY 1645 C-----CCAAAGGATGATGACCATGACATGAGCTGACCTTCTT 1677
Db 588 oSerSerGlyThrPheHisValLysLeuProLysArgArgGlyValGluLeuGlyIleTh 608
QY 1678 GCTGAAGTCC-----AAAAGGGAAACCAATTCACGCTTAGGAACACAGTGAGCT 1728
Db 608 rIleSerSerAlaSerArgLysArgGly-----617
QY 1729 CCGGCCCCACTCTGTGAACACAAAGCCTCGGACCGCTTGAGAGAGGCCACATGACACA 1788
Db 618 -GluProLeuIleIle-----SerAspIleLysLysGlySerValAlaH 632
QY 1789 CACC-----AGATGGCATCTTGGGACCTGAATCTATCACCAGGAATCTCAACATC-- 1840
Db 632 sArgThrGlyThrLeuGluProGlyAspLysLeuLeuAlaIleAspAsnIleArgLeuAs 652
QY 1841 -----CCTTTGGCC-----CTGAACACAGG 1860
Db 652 pHisCysProMetGluThrAlaValGlnIleLeuProGlnCysGluAspLeuValLysLe 672
QY 1861 CCAGATGAAGAACAGCTCGGCCACTTTTGTGAAGGCAATGTGGAGCA-----1909
Db 672 uLysIleArgLysAspGluAspAsnSerAspGluGlnGluSerSerGlyAlaValSerTy 692

QY 1910 -----AAGGAGCAGCCAGCCGTTTGGGA-----GAAGA 1938
Db 692 rThrValGluLeuLysArgTyrglyGlyProLeuGlyIleThrIleSerGlyThrGluG 712
QY 1939 TCtCAAGATCCAGACTCTCATTCCTTCTCTGCGCCAGTGAATTTGGTCTCTCCAGC 1998
Db 712 uProPheAspPro-IleIleIleSerGlyLeu-----ThrLysArgGlyLeuAlaGluA 730
QY 1999 TTTGGGAGACTCTCTTCCTGAACCTAATAAGACCCAGCTCTCTCTCTCCATC 2058
Db 730 rgThrGlyAlaSerMetLeuGlyThrAlaTyrrPrProSer***SerValSerLeuLysG 750
QY 2059 CTTCTCTCTGCTGCTCTCTTAATGCTGCCAGGATGTCTCAAACTTACTCTG 2117
Db 750 lyArgProLeuSerGluAlaIleHisLeuLeuGlnValAlaGlyGluThrValThrLeu 769
RESULT 11
US-09-045-632-2
; Sequence 2, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: HUGANIR, RICHARD L.
; APPLICANT: DONG, HUANG
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1112 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-045-632-2

Alignment Scores:
Pred. No.: 4,64e-09 Length: 1112
Score: 202.00 Matches: 144
Percent Similarity: 35.93% Conservatives: 82
Best Local Similarity: 22.89% Mismatches: 203
Query Match: 5.06% Indels: 201
DB: 3 Gaps: 26

US-09-502-945-5 (1-2162) x US-09-045-632-2 (1-1112)

QY 322 CAGCTGACCCCGCGCTCCAGGAAGCTGCGTGTGGACCGTGTGACCC 381

Db	42	GlaSerIleProGluGluPheLysGlySerThrValValGluLeuMetLysLysGluGly	61
Qy	382	GAAGGCGCTGGCTGAGTGGCTGGCTGGCTGGAGTTGGCTGCTGGCTCTTCATCTCC	441
Db	62	ThrThrLeuGlyCysThrValSerGlyGlyIleAspLysAspGlyLysProArgValSer	81
Qy	442	CACCTCATCAAGGCGGT---CAGGCACACAGCGTCGGGCTCCAGGTAGGAGCAGATC	498
Db	82	AsnLeuArgGlnGlyIleAlaAlaArgSerAspGlnLeuAspValGlyAspTyrIle	101
Qy	499	GTCCGGATCANTGGATATTCCATCTCCTCTGATCCCATAGGAGGTCAATCAACCTCAT	558
Db	102	LysAlaValAsnGlyIleAsnLeuAlaLysPheArgHisAspGluIleIleSerLeuLeu	121
Qy	559	CGA---ACCAAGAAAACCTGTGCTCCATCAAGTGACACATCGCGCTGATCCCGTGAAA	615
Db	122	LysAsnValGlyGluArgValValLeuGluValGluTyr-----GluLeuProVal	139
Qy	616	AGCTCTCCTGATGAGCCCTCACTTGGCAGTAGTGGATCAGTTGTGTGCGGAATCTGGG	675
Db	140	SerIleGlnGlySerSerValMetPheArgThrValGluValThrLeuHisLysGluGly	159
Qy	676	GGGTGGCAGGACCGTGTG-----GGTCCCTCGGAATCGGAAAACAGGAGAAGAAG	729
Db	160	AsnThrPheGlyPheValIleArgGlyGlyAlaHisAspAspArgAsnLysSerArgPro	179
Qy	730	GTCTTTCATCAGCTGTGTAGCTCCCGAGGCCCTTGGCTCGACATTCACAGCGGCCCATC	789
Db	180	ValValIleThr-----CysValArgProGlyGlyProAsp	191
Qy	790	CAGAAGCCTGGCATCTTTATCATGCCATGTGAACCTGGCTCCCTCTGCTGAGTGGGA	849
Db	192	AspArgGluGlyThr-----IleLysPro-----	199
Qy	850	TTGGAGATAGGGACCATGTCGAAGTATCGAGTGGCTGACTTCTCTAACTGGATCAC	909
Db	200	-----GlyAspArgLeuLeuSerValAspGlyIleArgLeuLeuGlyThrThrHis	216
Qy	910	AAGGAGGCTCTAATGTCTGAAA-----AATGCCGCGACCTGACCATC-----	954
Db	217	AlaGluAlaMetSerIleLeuLysGlnCysGlyGlnGluAlaThrLeuLeuIleGlyTyr	236
Qy	955	-----TCCATTGTAGCTGACGTGGCCCG-----GAGCTCTTTC	987
Db	237	AspValSerAlaMetAspSerValAlaThrAlaSerGlyProLeuLeuValGluValAla	256
Qy	988	ATGACACACCGGAGCGGTGGCAGAGCGCG-----	1019
Db	257	LysThr--ProGlyAlaSerLeuGlyValAlaLeuThrThrSerValCysAsnLysG	276
Qy	1020	-----GCAGCTGTAGCTGCAGCGCAGGACTTCTC	1050
Db	276	IleValIleValIleAspLysIleLysSerAlaSerIleAlaAspArgCysGlyAlaLeuH	296
Qy	1051	ATGCAGAGCGGTGGCGATGGAGTCCAAAGATCTCTCCAGGAGCAGCAGGATGGAG	1110
Db	296	is-----ValGlyAspHisIle-----LeuSerIleAspGlyT	307
Qy	1111	CGGCAAGAGGAGAAAATTTGCCCAAGAGGACAGGAGGAAAATTCAGAGATACCGGAAG	1170
Db	307	hrSerMetGluTyrCysThrLeuAlaGluAlaThrGlnPheLeu-----G	322
Qy	1171	GAGATGGAACAGATGTPAGAGGAGGAAGAGATTAAAGAACATGGGAAGAGATGG	1230
Db	322	LysThrThrAsp---GlnValLysLeuGluIleLeuProHisGlnThrArgLeuA	341
Qy	1231	GGCTCAAGGA-----	1241
Db	341	laLeuLysGlyProAspHisValLysIleGlnArgSerAspArgGlnLeuProItrasp	361
Qy	1242	-----ACAGCTACTTCTTGGCTTAAACCATCATCTGCTGAGGTACCCAGTA	1287

```

Db 361 roTPaLaSerSerGlnCysSerValHisThrAsnHisHis-----HisAsnProHisH 379
Qy 1288 CCCCTTCGCAAGCAAGTCAATTCGCGAAATATGAGGAAGGCTTTGACCCCT----- 1340
Db 379 isProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsnSerP 399
Qy 1341 -----ACTCTATGTTACCCAGAGCAGCATCATCGGGAAGGATGTCGCGCT 1386
Db 399 roProAlaMetValSerSerSerProThrSerMetSerAlaTyrSerLeuSerSerL 419
Qy 1387 C----- 1387
Db 419 euAsnMetGlyThrLeuProArgSerLeuTyrSerThrSerProArgGlyThrMetMetA 439
Qy 1388 --CTACGCATCAAGAAG-----GAGGGATCTCTAGACCTGGCC----- 1423
Db 439 rgArgArgLeuLysLysLysAspPheLysSerSerLeuSerLeuAlaSerSerThrValG 459
Qy 1423 ----- 1423
Db 459 lyLeuAlaGlyGlnValValHisThrGluThrThrGluValValLeuThrAlaAspProv 479
Qy 1424 -----CTGGAAGCGCGTGTG-----GACTCCTC 1445
Db 479 alThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThrGluThrLeuSerSerP 499
Qy 1446 CCATTGGGAAGGTGCTGTTCTGCTGCTATGACGCGGGAGCTCTGACGCGCATGGTG 1505
Db 499 ro-----ProLeuIleSerTyrIleGluAlaAspSerProAlaGluArgCysGlyV 516
Qy 1506 GCATTGTGAAGGGAGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTTACACCC 1565
Db 516 alLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIleProThrGluAspSerThrP 536
Qy 1566 TGGCTGAGGTGACGCTGCCCTCCAAAGGCCTGGNATCAGGCGGGGACCTGCATC--- 1621
Db 536 heGluGluAlaAsnGlnLeuLeuArgAspSerSerIleThrSerLysValThrLeuGluI 556
Qy 1622 -----GACCTTGTTGTTGCTGCTGCTGCCCT-----C 1646
Db 556 leGluPheAspValAlaGluSerValIleProSerSerGlyThrPheHisValLysLeuP 576
Qy 1647 CAAAGGAGTATGAGATGAGCTGACCTTCTGCTGAAGTCC----- 1687
Db 576 roLysLysHisSerValGluLeuGlyIleThrIleSerSerProSerSerArgLysProG 596
Qy 1688 -----AAAGGGGAACCAATTCACCGTTCAGGAACA 1721
Db 596 lyAspProLeuValIleSerAspIleLysGlySerValAlaHisArgThrGlyThrL 616
Qy 1722 GTGAGCTCCGCCCCACCTCGTG 1744
Db 616 euGluLeuGlyAspLysLeuLeu 623

RESULT 12
US-09-045-632-3
; Sequence 3, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Hugenair, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

RESULT 12

US-09-041-5632-3
 ? Sequence 3, Application US/090456323
 ?
 ? Patent No. 6001575
 ?
 ? GENERAL INFORMATION:
 ?
 ? APPLICANT: HUGANIR, Richard L.
 ? APPLICANT: Dong, Hualing
 ? TITLE OF INVENTION: THERAPEUTIC
 ? TITLE OF INVENTION: GRIP-RELATE
 ? NUMBER OF SEQUENCES: 105
 ?
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: DIKE, BRONSTEIN, R
 ? STREET: 130 Water Street
 ? CITY: Boston
 ? STATE: MA
 ? COUNTRY: USA
 ? ZIP: 02109
 ?
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ?


```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1112 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-045-632-3

Alignment Scores:
Pred. No.: 4,648-09 Length: 1112
Score: 202.00 Matches: 144
Percent Similarity: 35.93% Conservative: 82
Best Local Similarity: 22.89% Mismatches: 203
Query Match: 5.06% Indels: 201
DB: 3 Gaps: 26

US-09-502-945-5 (1-2162) x US-09-045-632-3 (1-1112)
QY 322 CAGCTGACCCCGCGCTCCAGAGAGTGAAGAGGTCGCTGACGCGTCTGCACCC 381
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
42 GlnSerIleProGluGluPheLysGlySerThrValValGluLeuMetLysLysGluGly 61
QY 382 GAAGCGCTCGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 441
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111
62 ThrThrLeuGlyCysThrValSerGlyGlyIleAspLysAspGlyLysProArgValSer 81
QY 442 CACCTCATCAAGCGGT---CAGCGACAGCGCTCGGCTCCAGTGGGACGAGATC 498
Db 111 111 111 111 111 111 111 111 111 111 111 111 111
82 AsnLeuArgGlnGlyIleAlaAlaArgSerGlnLeuAspValGlyAspTyrIle 101
QY 499 GTCCGGATCAATGGATATCCATCTCCTCTGATCCCATGAGGAGTGCATCACTCAT 558
Db 111 111 111 111 111 111 111 111 111 111 111 111 111
102 LysAlaValAsnGlyIleAsnLeuAlaLysPheArgHisAspGluIleSerLeuLeu 121
QY 559 CGA---ACCAAGAAACTGTCTCCAAAGTGACACACATCGGCTGATCCCGCGTGA 615
Db 111 111 111 111 111 111 111 111 111 111 111 111
122 LysAsnValGlyGluArgValValLeuGluValGluTyr-----GluLeuProVal 139
QY 616 ACCTCTCTGATGAGCCCTCACTGGCAGTATGGATCACTTTGTGCGAATCTGGG 675
Db 111 111 111 111 111 111 111 111 111 111 111 111
140 SerIleGlnGlySerSerValMetPheArgThrValGluValThrLeuHisLysGluGly 159
QY 676 GCGGTGCGAGGAGCGCTG---GGCTCCCTCGGAATCGGAAACAAAGAGAGAGAG 729
Db 111 111 111 111 111 111 111 111 111 111 111
160 AsnThrPheGlyPheValIleArgGlyGlyAlaHisAspArgAsnLysSerArgPro 179
QY 730 GTCCTTCATCAGCTGGTAGGCTCCGAGCCCTGGCTGCAGCATTTCCAGCGGCCCAT 789
Db 111 111 111 111 111 111 111 111 111 111 111
180 ValValIleThr-----CysValArgProGlyGlyProAsp 191
QY 790 CAGAAGCCTGGCATCTTTATCAGCATGTGAAACCTGGCTCCCTGCTGCTGAGTGGA 849
Db 111 111 111 111 111 111 111 111 111 111
192 AspArgGluGlyThr-----IleLysPro----- 199
QY 850 TTGGAGATAGGGGACGAGTTCGAATGCAATGCGGTGCGACTTCTCTAACCTGGATCAC 909

```

```

Db 200 -----GlyAspArgLeuLeuSerValAspGlyIleArgLeuLeuGlyThrThrHis 111
QY 910 AAGGAGGCTGTAATGTCTGAAA-----AATACCGCGAGCTCACCATC----- 954
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111
217 AlaGluAlaMetSerIleLeuLysGlnCysGlyGlnGluAlaThrLeuLeuIleGluTyr 236
QY 955 -----TCCATTGTAGCTGCAGCTGGCCGG-----CAGCTGTTC 987
Db 111 111 111 111 111 111 111 111 111 111 111 111 111
237 AspValSerAlaMetAspSerValAlaThrAlaSerGlyProLeuLeuValGluValAla 256
QY 988 ATGACAGACCGGCGGCTGGCAGCGCGG----- 1019
Db 111 111 111 111 111 111 111 111 111 111 111 111
257 LysThr--ProGlyAlaSerLeuGlyValAlaLeuThrThrSerValCysCysAsnLysG 276
QY 1020 -----GCAGCGTGTAGCTGCAGCGCGAGAGTTCCTC 1050
Db 111 111 111 111 111 111 111 111 111 111 111 111
276 InValIleValIleAspLysIleLysSerAlaSerIleAlaAspArgCysGlyAlaLeuH 296
QY 1051 ATGACAGACCGGCTGGCAGTCCCAACAAGATCTCCAGGAGCAGCAGGAGTGGAG 1110
Db 111 111 111 111 111 111 111 111 111 111 111 111
296 Is-----ValGlyAspHisIle-----LeuSerIleAspGlyT 307
QY 1111 CGGCAAGGAGAGAAATTCGCCAGAGCGCAGCAGAGGAAATGACAGATACCGGAG 1170
Db 111 111 111 111 111 111 111 111 111 111 111 111
307 hrSerMetGluTyrCysThrLeuAlaGluAlaThrGlnPheLeu-----G 322
QY 1171 GAGATGGAACAGATTGTAGAGGAGGAGAGAGTAAAGCAATGGAAGAGACTGG 1230
Db 111 111 111 111 111 111 111 111 111 111 111 111
322 LysnThrThrAsp---GlnValLysLeuGluIleLeuProHisGlnThrArgLeuA 341
QY 1231 GGTCAAAGGA----- 1241
Db 111 111 111 111 111 111 111 111 111 111 111 111
341 laLeuLysGlyProAspHisValLysIleGlnArgSerAspArgGlnLeuProTrpAsp 361
QY 1242 -----ACAGCTACTCTTGCCTAAACCATCACTGCTGAGTACACCCAGTA 1287
Db 111 111 111 111 111 111 111 111 111 111 111 111
361 roTrpAlaSerSerGlnCysSerValHisThrAsnHisHis-----HisAsnProHis 379
QY 1288 CCCCTCCCAAGCAAGTGTATTCGGGAAATATGAGGAGGCTTTGACCCCT----- 1340
Db 111 111 111 111 111 111 111 111 111 111 111 111
379 lsProAspHisCysArgValProAlaLeuGlyPheProLysAlaLeuThrProAsnSerP 399
QY 1341 -----ACTCTATGTTCAACCCAGCAGATCATGGGAGGAGTATGTCGGCT 1386
Db 111 111 111 111 111 111 111 111 111 111 111 111
399 roProAlaMetValSerSerSerProThrSerMetSerAlaTyrSerLeuSerSerL 419
QY 1387 C----- 1387
Db 111 111 111 111 111 111 111 111 111 111 111 111
419 euAsnMetGlyThrLeuProArgSerLeuTyrSerThrSerProArgGlyThrMetMetA 439
QY 1388 --TACGCGATCAAGAAG-----GAGGATCTCTTAGACCTGCGC----- 1423
Db 111 111 111 111 111 111 111 111 111 111 111 111
439 rgArgArgLeuLysLysLysAspPheLysSerSerLeuSerLeuAlaSerSerThrValG 459
QY 1423 -----CTGGAAGCGCGTGTG-----GACTCCC 1423
Db 111 111 111 111 111 111 111 111 111 111 111 111
459 lyLeuAlaGlyGlnValHisThrGluThrThrGluValValLeuThrAlaAspProv 479
QY 1424 -----CTGGAAGCGCGTGTG-----GACTCCC 1445
Db 111 111 111 111 111 111 111 111 111 111 111 111
479 alThrGlyPheGlyIleGlnLeuGlnGlySerValPheAlaThrGluThrLeuSerSerP 499
QY 1446 CCATTGGGAAGTGTGCTTCTGCTGTATGAGCGGGGAGCTGTGAGCGGATGGTG 1505
Db 111 111 111 111 111 111 111 111 111 111 111 111
499 ro-----ProLeuIleSerTyrIleGluAlaAspSerProAlaGluArgCysGlyV 516
QY 1506 GCATTGTGAAGGGGAGCAGATCATGGCAATCAACGGCAGATGTGTGACACTACACCC 1565
Db 111 111 111 111 111 111 111 111 111 111 111 111
516 alLeuGlnIleGlyAspArgValMetAlaIleAsnGlyIleProThrGluAspSerThrP 536
QY 1566 TGGCTGAGGCTGAGCTGCCCTCGCAAGAGCGCTGGAATCATAGGCGGGGACTGGATC 1621
Db 111 111 111 111 111 111 111 111 111 111 111 111

```


Db 536 heGluGluAlaAsnGlnLeuLeuArgAspSerSerIleThrSerLysValThrLeuGluI 556
QY 1622 -----GACCTTGTGGTCCGCTCCGCC-----C 1646
Db 556 leGluPheAspValAlaGluSerValIleProSerSerGlyThrPheHisValLysLeup 576
QY 1647 CAAAGAGTATCAGCATGACGTGACCTTCTGCTGAAGTCC----- 1687
Db 576 roLysLysHisSerValGluLeuGlyIleThrIleSerSerProSerSerArgLysProG 596
QY 1688 -----AAAAGGGAACCAATTCACCGGTTAGGAACA 1721
Db 596 lyAspProLeuValIleSerAspIleLysLysGlySerValAlaHisArgThrGlyThrL 616
QY 1722 GNGAGCTCGGGCCCACTCGNG 1744
Db 616 euGluLeuGlyAspLysLeuLeu 623

RESULT 13

US-09-045-632-14
; Sequence 14, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Huguair, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-045-632-14

Alignment Scores:

Pred. No.: 4e-09 Length: 604
Score: 201.00 Matches: 142
Percent Similarity: 36.47% Conservative: 83
Best Local Similarity: 23.01% Mismatches: 190
Query Match: 5.04% Indels: 203
DB: 3 Gaps: 27

US-09-502-945-5 (1-2162) x US-09-045-632-14 (1-604)

QY 364 CTGACCGTCTGCACCCCGAAGGC-----CTCGGCCTGAGTGTGCTGTGCTGGAG 417
Db 3 ValGluLeuMetLysLysGluGlyThrThrLeuGlyCysThrValSerGlyGlyLeuAsp 22
QY 418 TTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGCGCGT---CAGGCAGACAGCGTC 474
Db 23 LysAspGlyLysProArgValSerAsnLeuArgGlnGlyIleAlaAlaArgSerAsp 42
QY 475 GGGCTCCAGGTAGGGGACGAGATCGTCCGGATCAATGGATATTCATCTCTCTCTGTACC 534
Db 43 GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg 62
QY 535 CATGAGGAGTCAATCAACCTCATTCGA---ACCAAGAAACTGTGTCCATCAAGTGAAGA 591
Db 63 HisAspGluIleLeuSerLeuLeuLysAsnValGlyGluArgValValLeuGluValGlu 82
QY 592 CACATCGGCCTGATCCCGTGAAAGCTCTCTCTGATGAGCCCTCACTGGCAGTATGTG 651
Db 83 Tyr-----GluLeuProProValSerIleGlnGlySerSerValMetPheArgThrVal 100
QY 652 GATCAGTTTGTGTCGGAATCTGGGGCGTGGAGGACGCTG-----GGCTCCCTCGGA 705
Db 101 GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyGlyAlaHis 120
QY 706 AATCGGGAACAAAGAGAGAGAGGTCTTCATCAGCCTGTGTAGGTCCCGAGGCTTGGC 765
Db 121 AspAspArgAsnLysSerArgProValValIleThr----- 132
QY 766 TGCAGCATTTCCAGCGGCCCATCCAGAACCTGGCATCTTTATCAGCCATGTGAACCT 825
Db 133 CysValArgProGlyGlyProAspAspArgGluGlyThr-----IleLysPro 148
QY 826 GGCTCCCTGTCTGCTGAGGTGGATTGGAGATAGGGGACACAGATTGTGCAAGTCAATGGC 885
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157
QY 886 GTCGACTTCTTAACCTGGATCACAGGAGGCTGTAAATGTGCTGAA- - - - -AAT 936
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177
QY 937 AGCCGACGCTGACCATC-----TCCATTGTAGTGCAGCT 972
Db 178 GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer 197
QY 973 GGCCGG-----GAGCTTTTCATGACAGCGGGGAGCGGTGGCAGAGCGCG- - - - - 1019
Db 198 GlyProLeuLeuValGluValAlaLysThr--ProGlyAlaSerLeuGlyValAlaLeu 217
QY 1020 -----GCAGCGT 1026
Db 217 hrThrSerValCysAsnLysGlnValIleValIleAspLysLysSerAlaSerI 237
QY 1027 GAGCTGCAGCGGAGGAGCTTCTCATGCAGAAGCGGCTGGCGATGGATGCCAACAGATC 1086
Db 237 leAlaAspArgCysGlyAlaLeuHis-----ValGlyAspHisIle- - - - - 250
QY 1087 CTCAGGAGCAGCAGGAGATGGCGGCAAGAGAGAAAGAAATTCGCCAGAGACAGCA 1146
Db 251 -----LeuSerIleAspGlyThrSerMetGluTyrCysThrLeuAlaGluAlaThrG 268
QY 1147 GAGGAAATGAGAGATACCGGAAGAGAGATGGAACAGATGTGTAGAGGAGAGAGAGTTT 1206
Db 268 InPheLeu-----GlyAsnThrThrAsp---GlnValLysLeuGluIleL 282
QY 1207 AAGAAGCAATGGGAAGAACTGGGGCTCAAAGGA- - - - - 1241
Db 282 euProHisHisGlnThrArgLeuAlaLeuLysGlyProAspHisValLysIleGlnArgS 302
QY 1242 -----ACAGCTACTCTCTCCCTAAACACC 1263
Db 302 erAspArgGlnLeuProTrpAspProTrpAlaSerSerGlnCysSerValHisThrAsnH 322
QY 1264 ATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTGTATTCGGAATATGA 1323

Alignment Scores:		
Pred. No.:	4,36e-09	Length:
Score:	201.00	Matches:
Percent Similarity:	36.41%	Conservative:
Best Local Similarity:	23.01%	Mismatches:
Query Match:	5.04%	Indels:
DB:	3	Gaps:
		203
		27
		702

US-09-502-945-5 (1-2162) x US-09-045-632-15 (1-702)

QY	364	CTGCACGCTGTGCACCCCGAAGC-----CTCGGCTGAGTGTGGTGGCCTGGAG	417
Db	3	valgluLeuMetLysLysGluGlyThrLeuGlyCysThrValSerGlyGlyLeuAsp	22
QY	418	TTTGGCTGTGGCTCTTCATCTCCCACTCATCAAGGGCGT---CAGGCACACGCGTC	474
Db	23	LysAspGlyLysProArgValSerAsnLeuArgGlnGlyLeuAlaAlaArgSerAsp	42
QY	475	GGGCTCCAGCTAGGGACGAGATCGCCGGATCAATGGATATTCCTCTCCCTGTACC	534
Db	43	GlnLeuAspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg	62
QY	535	CATGAGGAGGTATCAACCTCATTCGA---ACCAGAAACTGTGTCCATCAAGTGAGA	591
Db	63	HisAspIuIleIleSerLeuLeuLysAsnValGlyGluArgValValLeuGluValGlu	82
QY	592	CACATCGGCCTGATCCCGCTGAAAGACTCTCCTGATGACGCCCTCACTTGGCAGTATGTG	651
Db	83	Tyr-----GluLeuProValSerIleGlnGlySerValMetPheArgThrVal	100
QY	652	GATCAGTTTGTGTGCGAATCTGGGCGGTGCGAGCAGCGCTG-----GGCTCCCTCGGA	705
Db	101	GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyGlyAlaHis	120
QY	706	AATCGGGAAACAGGAGAAGGTCTTCATCAGCCTGGTAGGCTCCGAGGCCTTGGC	765
Db	121	AspAspArgAsnLysSerArgProValValIleThr-----	132
QY	766	TGCAGCATTTCCAGCGCGCCCATCCAGAGCCTGGCATCTTTATCAGCCATCTGAACCT	825
Db	133	CysValArgProGlyGlyProAspArgGluGlyThr-----IleLysPro	148
QY	826	GGCTCCTGTCTGCTGAGTGGATTTGGAGATAGGAGGCACGATTGTCGAAGTCAATGCG	885
Db	149	-----GlyAspArgLeuLeuSerValAspGly	157
QY	886	GTGCACTTCTCAACCTGGATCACAGGAGGCTGTAAATGTGCTGAAA-----AAT	936
Db	158	IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln	177

SOFTWARE: PatentIn Release #1.0, Version #1.30

Qy	937	AGCCGAGCGCTGACCATC-----TCCATTGTAGCTGCAGCT	972
Db	178	GlulAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer	197
Qy	973	GGCGG-----CAGCTGTTTCATGACAGACCGGAGCGCTGCAGAGCGCG---	1019
Db	198	GlyProLeuLeuValGluValAlaLysThr--ProGlyAlaSerLeuGlyValAlaLeuT	217
Qy	1020	-----GCACGCGT	1026
Db	217	hrThrSerValCysCysAsnLysGlnValIleValIleAspLysIleLysSerAlaSerI	237
Qy	1027	GAGCTGCGCGGAGAGCTTCTCATGCGAAGCGGTGGCGATGGAGTCCACACAGATC	1086
Db	237	leAlaAspArgCysGlyAlaLeuHis-----ValGlyAspHisIle---	250
Qy	1087	CTCCAGGAGCAGCAGGAGATGGAGCGGCAAGGAGAAAGAAATGTCGCCAAGAGCAGCA	1146
Db	251	-----LeuSerIleAspGlyThrSerMetGluTyrCysThrLeuAlaGluAlaThrG	268
Qy	1147	GAGGAAATGAGAGATACCGGAAGAGGATGGACAGATGTAGAGGAGGAAGAGATTT	1206
Db	268	lnPheLeu-----GlyAsnThrThrAsp---GlnValLysLeuGluIleL	282
Qy	1207	AAGRAGCAATGGGAAGAGACTGGCGCTCAAAGGA-----	1241
Db	282	euProHisHisGlnThrArgLeuAlaLeuLysGlyProAspHisValLysIleGlnArgS	302
Qy	1242	-----ACAGCTACTCTGCTCCCTAAAACC	1263
Db	302	erAspArgGlnLeuProTrpAspProTrpAlaSerSerGlnCysSerValHisThrAsnH	322
Qy	1264	ATCACTGCTGAGGTACACCCGATCCCTCCGACGCCAAAGTGTATTCGGAAATATGA	1323
Db	322	IsHis-----HisAsnProHisHisProAspHisCysArgValProAlaLeuGlyPheP	340
Qy	1324	GGAAGGCTTTCACCCCT-----ACCTATGTTCCACCCAGCAGCA	1362
Db	340	roLysAlaLeuThrProAsnSerProProAlaMetValSerSerSerProThrSerM	360
Qy	1363	GATCATGGGAAGGATGTCGGGCTC-----	1387
Db	360	etSerAlaTyrSerLeuSerSerLeuAsnMetClyThrLeuProArgSerLeuTyrSerT	380
Qy	1388	-----CTACGCATCAAGAAG-----GAGGGATCCT	1412
Db	380	hrSerProArgGlyThrMetMetArgArgLeuLysLysLysAspPheLysSerSerL	400
Qy	1413	TAGACCTGGCC-----CTGGAAGCGCGTGTG-	1423
Db	400	euSerLeuAlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrThrG	420
Qy	1424	-----CTGGAAGCGCGTGTG-	1438
Db	420	luValValLeuThrAlaAspProValThrClyPheGlyIleGlnLeuGlnGlySerValP	440
Qy	1439	-----GACTCCCCCATTTGGGAAGGTGGTCTCTTCTGCTGCTATGACG	1481
Db	440	heAlaThrGluThrLeuSerSerPro-----ProLeuIleSerTyrIleGluAlaA	457
Qy	1482	GGGAGCTGCTGAGCGGCATGTGGCATTTGTAAGGGGACGAGATCATGCGCATCAACG	1541
Db	457	spSerProAlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetalIleAsnG	477
Qy	1542	GCAAGATTGTACAGACTACACCTGGCTGAGGCTGACGCTGCCCTGCAGAGGCGCTGGA	1601
Db	477	lyIleProThrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerSerI	497
Qy	1602	ATCAGGCGGGGACTGGATC-----GACCTTGTGGTGGCGGCTGCCCC-----	1645
Db	497	leThrSerLysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerS	517
Qy	1646	-----CCAAAGGAGTATGACGATGACGATGACGCTTCTTCTGTGA	1682

```

Db 517 erGlyThrPheHisValLysLeuProLysLysHisSerValGluLeuGlyIleThrIles 537
      |||||..... |||||
QY 1683 AGTCC-----AAAAGGGGAA 1697
      |||
Db 537 erSerProSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlyS 557
      |||
QY 1698 ACCAAATTCACGGTTAGAAACAGTGAGCTCCGGGCCCCACTCGTG 1744
      :||| ||| |||||
Db 557 erValAlaHisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572
      :||| ||| |||||

RESULT 15
US-09-045-632-16
; Sequence 16, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Huananr, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-045-632-16

Alignment Scores:
Pred. No.: 5.4e-09 Length: 1018
Score: 201.00 Matches: 142
Percent Similarity: 36.47% Conservative: 83
Best Local Similarity: 23.01% Mismatches: 190
Query Match: 5.04% Indels: 203
DB: 3 Gaps: 27

```

Alignment Scores:	
Pred. No.:	5.4e-09
Score:	201.00
Percent Similarity:	36.4%
Best Local Similarity:	23.01%
Query Match:	5.04%
OP:	3
Length:	1018
Matches:	142
Conservative:	83
Mismatches:	190
Indels:	203
Gaps:	27

US-09-502-945-5 (1-2162) x US-09-045-632-16 (1-1018)

364	CTGACCGCTGACACCCGGAAGG-----CTGGCGCTGAGTGTGGTGGTGGCTGGAG	417
QY		
3	ValGluLeuMetLysLysGluGlyThrThrLeuGlycysThrValSerGlyGlyIleasp	22
DB		
418	TTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAAAGGGCGT---CAGGCACACGCGTC	474
QY		
23	LysAspGlyLysProArgValSerAsnLeuArgGlnGlyGlyIleAlaalaArgSerAsp	42
DB		


```
QY 475 GGCCTCCAGTAGGAGGAGATCGTCCGGATCAATGGATATTCTCATCTCCTCTGTAACC 534
    ||| ||||||| ||| ::::| ||| ::::| ||| ::::| |||
Db 43 GlnLeuaspValGlyAspTyrIleLysAlaValAsnGlyIleAsnLeuAlaLysPheArg 62

QY 535 CATGAGGAGGTCAATCAACCTCATTCGA---ACCAAGAAACTGTGTCCATCAAGTGA 591
    ||| ||||||| ||| ::::| ||| ::::| ||| ::::| |||
Db 63 HisAspGluIleSerLeuLeuLysAsnValGlyGluArgValValLeuGluValGlu 82

QY 592 CACATCGGCTGTATCCCGTGAAGACTCTCTGTAGTACGCCCTCACTTGGCAGTATG 651
    ::::| ||| ||| ::::| ||| ::::| |||
Db 83 Tyr-----GlnLeuProProValSerIleGlnGlySerValMetPheArgThrVal 100

QY 652 GATCACTTGTGTGCGAATCTGGGCGTCCGAGGAGCAGCTG-----GCCTCCCTGGA 705
    ::::| ||| ||| ::::| |||
Db 101 GluValThrLeuHisLysGluGlyAsnThrPheGlyPheValIleArgGlyGlyAlaHis 120

QY 706 ATTCGGGAAACAAGAGAGAGGTCTTCATCAGCAGCTGTGTATGATCTTATCAGCCATGTGAACCT 825
    ::::| ||| ||| ::::| |||
Db 121 AspAspArgAsnLysSerArgProValValIleThr----- 132

QY 766 TGCAGCATTTCCAGCGGCCCATCCAGAGCCTGGCATCTTTATCAGCCATGTGAACCT 825
    ||| ||||||| ||| ::::| |||
Db 133 CysValArgProGlyGlyProAspAspArgGluGlyThr-----IleLysPro 148

QY 826 GGCTCCCTCTGCTGCTGAGTGGGATTGGAGATAGGGGACACAGATTCTCGAAGTCAATGCG 885
    ||| ||||||| ||| ::::| |||
Db 149 -----GlyAspArgLeuLeuSerValAspGly 157

QY 886 GTGCACTTCTTAACCTGGATCACAGGAGGTGTAATGTGCTGAAA-----AAT 936
    ::::| ||| ||| ::::| |||
Db 158 IleArgLeuLeuGlyThrThrHisAlaGluAlaMetSerIleLeuLysGlnCysGlyGln 177

QY 937 ACCCGCAGCTGACCATC-----TCCATTGTAGCTGCAGCT 972
    ::::| ||| ||| ::::| |||
Db 178 GluAlaThrLeuLeuIleGluTyrAspValSerAlaMetAspSerValAlaThrAlaSer 197

QY 973 GGCCGG-----GAGCTGTTTCATGACAGCCGCGGCGTGGCAGAGCGCG--- 1019
    ||| ||||||| ||| ::::| |||
Db 198 GlyProLeuLeuValGluValAlaLysThr--ProGlyAlaSerLeuGlyValAlaLeu 217

QY 1020 -----GCAGCGT 1026
    ||| ||||||| |||
Db 217 hrThrSerValCysAsnLysGlnValIleValIleAspLysIleLysSerAlaSerI 237

QY 1027 GAGCTCCAGCGCAGGAGTCTTCATGACAGAGCGCTGCGATGGATGCCAACAGATC 1086
    ||| ||||||| ||| ::::| |||
Db 237 leAlaAspArgCysGlyAlaLeuHis-----ValGlyAspHisIle----- 250

QY 1087 CTCAGGAGCAGGAGATGGAGCGGCAAGAGGAAAGAAATTCGCCAGAGGAGCGCA 1146
    ::::| ||| ||| ::::| |||
Db 251 -----LeuSerIleAspGlyThrSerMetGluTyrCysThrLeuAlaGluAlaThrG 268

QY 1147 GAGGAAATGAGAGATACCGGAGGAGATGGAACAGATTGTAGAGGAGGAGAGAGTTT 1206
    ::::| ||| ||| ::::| |||
Db 268 InPheLeu-----GlyAsnThrThrAsp--GlnValLysLeuGluIleL 282

QY 1207 AAGAGCAATGGGAAGACACTGGGCTCAAGGA----- 1241
    ||| ||||||| |||
Db 282 euProHisHisGlnThrArgLeuAlaLeuLysGlyProAspHisValLysIleGlnArgS 302

QY 1242 -----ACAGCTACTCTTGCTTAAACCC 1263
    ::::| ||| ||| ::::| |||
Db 302 erAspArgGlnLeuProTrrpAspProTrrpAlaSerSerGlnCysSerValHisThrAsnH 322

QY 1264 ATCACTGTGAGGTACACCCAGTACCCCTTCGCAAGCCCAAGTATTTCCGGAAATATGA 1323
    ||| ||| ||| ::::| |||
Db 322 ishHis-----HisAsnProHisHisProAspHisCysArgValProAlaLeuGlyPheP 340

QY 1324 GGAAGCTTTGACCCCT-----ACTATGTTTACCCCGAGCA 1362
    ||| ||||||| |||
Db 340 roLysAlaLeuThrProAsnSerProProAlaMetValSerSerSerProThrSerM 360
```

```
QY 1363 GATCATGGGAAGGATGTCCGGCTC----- 1387
    ||| ::::| |||
Db 360 etSerAlaTyrSerLeuSerSerLeuAsnMetGlyThrLeuProArgSerLeuTyrSer 380

QY 1388 -----CTACGCATCAAGAAG-----GAGGATCCT 1412
    ||| ||||||| ||| ::::| |||
Db 380 hrSerProArgGlyThrMetMetArgArgGluLeuLysLysLysAspPheLysSerSerL 400

QY 1413 TAGACCTGGCC----- 1423
    || |||||||
Db 400 euSerLeuAlaSerSerThrValGlyLeuAlaGlyGlnValValHisThrGluThrThrG 420

QY 1424 -----CTGGAAGCGGTGTG- 1438
    ||| ||||||| |||
Db 420 luValValLeuThrAlaAspProValThrGlyPheGlyIleGlnLeuGlnGlySerValP 440

QY 1439 -----GACTCCCCCATGGGAAGGTGGTCTTCTGCTGTGTATGAGC 1481
    ||||||| |||
Db 440 heAlaThrGluThrLeuSerSerPro-----ProLeuIleSerTyrIleGluAlaA 457

QY 1482 GGGGAGCTGCTGAGCGGCATGGTGCATTTGTAAAGGGGAGAGATCATGCAATCAACG 1541
    ||||||| ||| ::::| |||
Db 457 spSerProAlaGluArgCysGlyValLeuGlnIleGlyAspArgValMetAlaIleAsnG 477

QY 1542 GCAAGATTGTGACACACTACACCTGGCTGAGGCTGAGGCTGCCCTGCAGAGGCTCGA 1601
    || ||| ||| ||| |||
Db 477 lyIleProThrGluAspSerThrPheGluGluAlaAsnGlnLeuLeuArgAspSerSerI 497

QY 1602 ATCAGGGCGGGACGTGATC-----GACCTTGTGGTTCCTGCTGCCCTC----- 1645
    ::::| ||| ||| ::::| |||
Db 497 leThrSerLysValThrLeuGluIleGluPheAspValAlaGluSerValIleProSerS 517

QY 1646 -----CCAAAGGAGTATGACGATGACGTGACCTTCTTGTGCTGA 1682
    ||| ||||||| |||
Db 517 erGlyThrPheHisValLysLeuProLysLysHisSerValGluLeuGlyIleThrIleS 537

QY 1683 AGTCC-----AAAAGGGGAA 1697
    ||| |||
Db 537 erSerProSerSerArgLysProGlyAspProLeuValIleSerAspIleLysLysGlyS 557

QY 1698 ACCAAATTCACGCGTTAGGAACACAGTGCAGCTCCGCCGCCCTCGT 1744
    :: ||| ||| |||
Db 557 erValAlaHisArgThrGlyThrLeuGluLeuGlyAspLysLeuLeu 572
```

Search completed: March 21, 2003, 13:09:23
Job time : 57.7489 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 13:00:02 ; Search time 23.1498 Seconds
(without alignments)
9985.861 Million cell updates/sec

Title: US-09-502-945-5
Perfect score: 3989
Sequence: 1 cctgcccgcgcgcgtcgc.....tattttccagcttaaaaaa 2162

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 442306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_6/ptodata/US09502945/runat_14032003_101101_19301/app_query.fasta_1.10979
-DB=PublishedApplications_AA -OFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09502945 -CGN_1_1_60 -runat_14032003_101101_19301
-NCPU=6 -ICPU=3 -NO_XLPUX -NO_MMAL -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PublishedApplications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PTC_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PTC_US_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	1175.5	29.5	268	10 US-09-739-907-114
2	435	10.9	120	9 US-09-925-299-778
3	435	10.9	120	10 US-09-925-299-778
4	417	10.5	98	10 US-09-739-907-60

5	417	10.5	113	10	US-09-739-907-119	Sequence 119, App
6	406	10.2	95	10	US-09-739-907-111	Sequence 111, App
7	318.5	8.0	260	9	US-10-078-090-151	Sequence 151, App
8	263	6.6	2037	9	US-09-951-402-3	Sequence 3, Appli
9	263	6.6	2037	10	US-09-951-401-3	Sequence 3, Appli
10	263	6.6	2037	10	US-09-922-101-3	Sequence 3, Appli
11	250.5	6.3	1881	9	US-09-998-425-3	Sequence 3, Appli
12	250.5	6.3	1881	9	US-09-997-977-3	Sequence 3, Appli
13	237	5.9	46	10	US-09-864-761-39112	Sequence 39112, A
14	234.5	5.9	767	10	US-09-919-497-59	Sequence 59, Appl
15	232	5.8	45	10	US-09-864-761-39113	Sequence 39113, A
16	225	5.7	47	10	US-09-864-761-49113	Sequence 49113, A
17	225	5.6	1736	10	US-09-919-497-98	Sequence 98, Appl
18	204.5	5.1	2485	10	US-09-802-669-46	Sequence 46, Appl
19	194	4.9	744	10	US-09-862-179A-1	Sequence 1, Appli
20	190	4.8	1337	10	US-09-757-781-62	Sequence 62, Appl
21	184	4.6	358	9	US-10-274-968-4	Sequence 4, Appli
22	184	4.6	358	10	US-09-823-356-6	Sequence 6, Appli
23	184	4.6	358	10	US-09-740-027-4	Sequence 4, Appli
24	183.5	4.6	1356	10	US-09-757-781-2	Sequence 2, Appli
25	181	4.5	434	12	US-10-001-843-179	Sequence 179, App
26	178	4.5	632	9	US-09-992-598-219	Sequence 219, App
27	178	4.5	632	9	US-09-989-293A-219	Sequence 219, App
28	178	4.5	632	9	US-10-063-547-40	Sequence 40, Appl
29	178	4.5	632	9	US-09-989-735-219	Sequence 219, App
30	178	4.5	632	9	US-09-990-444-219	Sequence 219, App
31	178	4.5	632	9	US-09-989-730-219	Sequence 219, App
32	178	4.5	632	9	US-09-990-436-219	Sequence 219, App
33	178	4.5	632	9	US-09-991-181-219	Sequence 219, App
34	178	4.5	632	9	US-09-993-687-219	Sequence 219, App
35	178	4.5	632	9	US-09-989-734-219	Sequence 219, App
36	178	4.5	632	9	US-09-997-653-219	Sequence 219, App
37	178	4.5	632	9	US-10-174-590-166	Sequence 166, App
38	178	4.5	632	9	US-10-176-758-166	Sequence 166, App
39	178	4.5	632	9	US-10-063-616-40	Sequence 40, Appl
40	178	4.5	632	9	US-10-175-737-166	Sequence 166, App
41	178	4.5	632	9	US-09-993-667-219	Sequence 219, App
42	178	4.5	632	9	US-10-083-502-40	Sequence 40, Appl
43	178	4.5	632	9	US-10-173-706-166	Sequence 166, App
44	178	4.5	632	9	US-10-175-738-166	Sequence 166, App
45	178	4.5	632	9	US-10-175-752-166	Sequence 166, App

ALIGNMENTS

RESULT 1
US-09-739-907-114
; Sequence 114, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 114
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-907-114


```
Alignment Scores:
Pred. No.: 4,75e-81 Length: 268
Score: 1175.50 Matches: 236
Percent Similarity: 90.11% Conservative: 1
Best Local Similarity: 89.73% Mismatches: 0
Query Match: 29.47% Indels: 26
DB: 10 Gaps: 1

US-09-502-945-5 (1-2162) x US-09-739-907-114 (1-268)
QY 964 GCTGAGCTGCGCGGAGCTGTTCTATGACAGACCGGGAGCGCTGGCAGAGCGCGGCAG 1023
DB 6 AAlaAlaGlyArgGluLeuPheMetThrAspArgGluArgLeuAlaGluAlaArgIn 25
QY 1024 CQTGAGCTGACGCGCAGAGCTTCTCATGCAAGAGCGCTGGCGCATGGATGCCAACAG 1083
DB 26 ArgGluLeuGlnArgGlnGluLeuMetGlnLysArgLeuAlaMetGlnSerAsnLys 45
QY 1084 ATCTCCAGGACGACGAGATGGAGCGGCAAGAGAGAAATGCCCAGAGGCA 1143
DB 46 IleLeuGlnGlnGlnGlnMetGluArgGlnArgGlnLysGluIleAlaGlnLysAla 65
QY 1144 GCAGAGAAATGAGATACCGGAGGAGATGGACAGATTGTAGAGGAGGAGAGAG 1203
DB 66 AlaGluGluAsnGluArgTyrArgLysGluMetGluGlnIleValGluGluGluLys 85
QY 1204 TTTAAGAACAATGGGAAGACACTGGGGCTCAAAGGACACGCTACTCTTGCTTAAACC 1263
DB 86 PheLysLysGlnTrpGluLysPrpGlySerLysGluGlnLeuLeuLeuProLysThr 105
QY 1264 ATCACTGCTAGGTACACCCAGTACCCCTTCGCAAGCCA----- 1302
DB 106 IleThrAlaGluValHisProValProLeuArgLysProLysTyrAspGlnGlyValGlu 125
QY 1303 -----AAGTGA 1308
DB 126 ProGluLeuGluProAlaAspLeuAspGlyGlyThrGluGluGlnGlyGln-As 145
QY 1309 TTTCCGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTTTCAACCCAGACAGATCAT 1368
DB 145 PheArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGluGlnIleMe 165
QY 1369 GGGGAAGATGTCGGCTCCTACGCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGA 1428
DB 165 tGlyLysAspValArgLeuLeuArgIleLysLysGluGlySerLeuAspLeuAlaLeuG1 185
QY 1429 AGCGGTGTGGACTCCCTCCATTTGGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1488
DB 185 uGlyGlyValAspSerProIleGlyLysValValSerAlaValTyrGluArgGlyAl 205
QY 1489 TGCTGACGCGCATGTGTGGCATTGTGAAGGGGACGAGATCATGGCAATCAACGGCAAGAT 1548
DB 205 aAlaGluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLysIl 225
QY 1549 TGTGACAGACTACACCTCGCTGAGGCTGACGCTGCGCTGCGAGAGCGCTGGAATCAGGG 1608
DB 225 eValThrAspTyrThrLeuAlaGluAlaAspAlaAlaLeuGlnLysAlaTrpAsnGlnG1 245
QY 1609 CGGGGACTGGATCGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1668
DB 245 yGlyAspTrpIleAspLeuValValAlaValCysProProLysGluTyrAspAspGluLe 265
QY 1669 GACCTTC 1675
DB 265 uThrPhe 267

RESULT 2
US-09-925-299-778
; Sequence 778, Application US/09925299
; Patent No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL02
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0

US-09-502-945-5 (1-2162) x US-09-925-299-778 (1-120)
QY 1196 AAGAGAGTTTAAAG---AGCAATGGGAAGAGACTGGGCTCAAAGAGACAGTACTCT 1252
DB 6 GluArgHisLeuArgValThrGlnGlyLeuGlyThrGlyAlaPheLeuGlyGly----- 23
QY 1253 TGCCTAAACCACTCACCTGCTGAGCTACACCCAGTACCCCTTCGCAAGCCAAAGTGATTTC 1312
DB 24 ---LeuArgProValLeu-----GlnProArgGlnGlyGln---AspPhe 36
QY 1313 CGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTTTACCCAGAGACAGATCATGGGG 1372
DB 37 ArgLysTyrGluGluGlyPheAspProTyrSerMetPheThrProGluGlnIleMetGly 56
QY 1373 AAGATGTCCGGCTCCCTACGCATCAAGAGGAGGATCCTTAGACCTGGCCCTGGAAGGC 1432
DB 57 LysAspValArgLeuLeuArgIleLysGluGlySerLeuAspLeuAlaLeuGluGly 76
QY 1433 GGTGTGGACTCCCATATTGGGAAGGTGCTCTTCTGCTGTGTATGAGCGGGAGCTGCT 1492
DB 77 GlyValAspSer**IleGlyLysValValSerAlaValTyrGluArgGlyAlaAla 96
QY 1493 GAGCGCATGTGGCATTTGAAAGGGGACGAGATCATGGCAATCAACGGCAAGTGTG 1552
DB 97 GluArgHisGlyGlyIleValLysGlyAspGluIleMetAlaIleAsnGlyLysIleVal 116
QY 1553 ACAGACTACACC 1564
DB 117 ThrAspTyrThr 120

RESULT 3
US-09-925-299-778
; Sequence 778, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL02
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2001-08-10
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
```



```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (98)
; OTHER INFORMATION: xaa equals stop translation
US-09-739-907-60

Alignment Scores:
Pred. No.: 1,04e-23 Length: 88
Score: 417.00 Matches: 88
Percent Similarity: 83.81% Conservative: 0
Best Local Similarity: 83.81% Mismatches: 0
Query Match: 10.45% Indels: 17
DB: 1 Gaps: 1

US-09-502-945-5 (1-2162) x US-09-739-907-60 (1-98)
QY 676 GGGTGGGAGCGACCTGGCTCCCTCGGAATCGGAAACAAGGAGAGAGGTCCTTC 735
|||
Db 10 GlyValArgGlySerLeuGlySerProGlyAsnArgGluAsnLysGlyLysValPhe 29
|||
QY 736 ATCAGCCTGTAGGCTCCGAGGCGCTTGCTGCAGCATTTCCAGCGCGCCCATCCAGAAG 795
|||
Db 30 IleSerLeuValGlySerArgGlyLeuGlyCysSerIleSerSerGlyProIleGlnLys 49
|||
QY 796 CCTGGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGCTGCTGAGGTGGGATTGGAG 855
|||
Db 50 ProGlyIlePheIleSerHisValLysProGlySerLeuSerAlaGluValGlyLeuGlu 69
|||
QY 856 ATAGGGAGCAGATTGTCGAAGTCATATGGCTGCAGCTTCTCTAACTGGATCACAAGGAG 915
|||
Db 70 IleGlyAspGlnIleValGluValAlaAnglyValAspPheSerAsnLeuAspHisLysGlu 89
|||
QY 916 GCTGTAAATGCTGTAATAATAGCCGAGCCTGACCATCTCCATTGTAGCTGCAGCTGGC 975
|||
Db 90 -Leu-----GlnLeuAl 93
|||

QY 976 CGGAGAGCTGTCA 988
|||
Db 93 aGlySerCysSer 97

RESULT 5
US-09-739-907-119
; Sequence 119, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: PZ022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 119
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-907-119

Alignment Scores:
Pred. No.: 1.11e-23 Length: 113
Score: 417.00 Matches: 88
Percent Similarity: 83.81% Conservative: 0

```


Best Local Similarity: 83.81% Mismatches: 0
Query Match: 10.45% Indels: 17
DB: 10 Gaps: 1

US-09-502-945-5 (1-2162) x US-09-739-907-119 (1-113)

Qy 676 GCGTGGCAGGAGCGCTGGCTCCCTGGAATCGGAAACAAAGGAGAAAGGTCTTC 735
Db 26 GlyValArgGlySerLeuGlySerProGlyAsnArgGluAsnLysGluLysValPhe 45
Qy 736 ATCAGCCTGGTAGGCTCCCGAGGCGCTGGCTGCAGCATTTCCAGGGCCCATCCAGAG 795
Db 46 IleSerLeuValGlySerArgGlyLeuGlyCysSerIleSerSerGlyProIleGlnLys 65
Qy 796 CTTGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGCTGCTGAGTGGGATTGGAG 855
Db 66 ProGlyIlePheIleSerHisValLysProGlySerLeuSerAlaGluValGlyLeuGlu 85
Qy 856 ATAGGGGACCAAGATTGTGCAAGTCAATGGCGTGCATCTCTTAACCTGGATCAACAAGGAG 915
Db 86 IleGlyAspGlnIleValGluValAsnGlyValAspPheSerAsnLeuAspHisLysGlu 105
Qy 916 GCTGTAAATGTCTGCAAAATAGCGCAGCGCTGACCATCTCCATTGTAGCTGCAGCTGCC 975
Db 106 -Leu-----GlnLeuAl 109

Qy 976 CGGAGCTGTTC 988

Db 109 aclySerCysSer 113

RESULT 6

US-09-739-907-111
; Sequence 111, Application US/09739907
; Patent No. US2001001289A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 111
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-907-111

Alignment Scores:
Pred. No.: 7 02e-23 Length: 95
Score: 406.00 Matches: 90
Percent Similarity: 60.13% Conservative: 2
Best Local Similarity: 58.82% Mismatches: 1
Query Match: 10.18% Indels: 60
DB: 10 Gaps: 1

US-09-502-945-5 (1-2162) x US-09-739-907-111 (1-95)

Qy 1678 GCTGAAGTCAAAAGGGGAAACCAATTCACGCGTTAGGAACAGTAGCTCGGCCCA 1737
Db 3 SerGluValGlnLysGlyLysProAsnSer-ProLeuGlyAsnSerGluLeuArgProHi 22

Qy 1738 CCTCGTGAACACAAAGCCCTCGGACCGCTTGAGAGAGCCACATGACACACACAGATG 1797

Db 22 sLeuValAsnThrLysProArgThrSerLeuGluArgGlyHis----- 36
Qy 1798 GCATCCTTGGGACCTGAATCTATCACCAGGAATCTCAACTCCCTTTGGCCCTGAACCA 1857
Db 36 ----- 36
Qy 1858 GGGCCAGATAAGGAACAGCTCGGGCCACTTTTTTTGAAGGCCAATGTGGAGAAAGGAGC 1917
Db 36 ----- 36
Qy 1918 AGCCAGCCGTTTGGGAGAGAGATCTCAAGGATCCAGACTCTCATCTCTTCTCTGGCCCA 1977
Db 37 -----Thr-IleProPheLeuTrippros 44
Qy 1978 GTGAATTTGGTCTCTCCAGCTTTGGGGGACTCCTCTTTGAACCCCTTAATAAGACCCAC 2037
Db 44 erGluPheGlyLeuSerGlnLeuTrpGlyThrProSerLeuAsnProAsnLysThrProL 64
Qy 2038 TGGAGTCTCTCTCTCCATCCCTCTCTCTGCGCTCTGCTCTAAATGCTGCCAGGATTG 2097
Db 64 euGluSerLeuSerLeuHisProSerProLeuProSerAlaLeuIleAlaAlaArgilev 84
Qy 2098 TCACCTCCAAACCTTACTCTGAGCTCATTAATAAA 2132
Db 84 alThrProAsnLeuThrLeuSerSerLeuIleLys 95

RESULT 7

US-10-078-090-151
; Sequence 151, Application US/10078090
; Publication No. US20030044815A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karra, Kalpana
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and
; FILE REFERENCE: DEX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 151
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-078-090-151

Alignment Scores:
Pred. No.: 4 64e-16 Length: 260
Score: 318.50 Matches: 85
Percent Similarity: 52.32% Conservative: 39
Best Local Similarity: 35.86% Mismatches: 84
Query Match: 7.98% Indels: 29
DB: 9 Gaps: 6

US-09-502-945-5 (1-2162) x US-10-078-090-151 (1-260)

Qy 358 GTGCTCTGGACCGTCTGCACCCCGAGGCGCTGAGTGTGCGTGGTGGCTGGAG 417
Db 1 LeuArgArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleArgGlyGlySerGlu 19
Qy 418 TTGCGTGTGGGCTCTTCATCTCCACCTCATCAAGCGCGTCAAGCAGACAGCGTGGG 477
Db 20 HisGlyValGlyIleTyValserLeuValGluProGlySerLeuAlaGluLysGluGly 39

Qy 478 CTCAGGTAGGGGACGAGATCGTCCGGATCAATGGATATTCCTCTCTCTCTGATCCCAT 537


```

Db 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 538 GAGGAGGTGATCAACCTCATTCAGAACCAAGAACTGTGTCCATCAAACTGAGACACATC 597
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysValLeuValLysSerAla 79
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 598 GGCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTCCTCTGAGGAGATGTGAT--- 554
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAspPro 97
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 655 -----CAGTTTGTGCGAATCTGGGGCTCGGAGCGAGCGCTGGC----- 696
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 98 GlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyAlaLeuArg 117
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 697 ---TCCCTGGAAATCGGAAAC-----AAGGAGAGAAAG 729
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 118 GlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyAspGluLysLys 137
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 730 GTCTTCATCAGCTGCTAGCTCCGAGCGCTGGCTGAGCATTTCCAGCGCGCCATC 789
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 790 CAGAGAGCTGGCATCTTTATCAGCATGTGAACCTGGCTCCCTGCTCTGAGTGGA 849
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 158 TyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly 177
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 850 TTGGAGATAGGAGGACAGATTCGAGTCAATGGCGTGCAGTCTCTTAACCTGGATCAC 909
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 910 AAGGAGGCTGTAATGTGCTGAAATACCCAGCGCTGACCATCTCCATTCTAGCTGCA 969
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThrVal----- 214
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 970 GCTGGCGGAGCTGTTTCATGACAGCGGAGCGGCTGGCAGAGCGCGG 1020
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 215 -----LysaspvalGlyArgLeuProHisAlaArg 224
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 8
US-09-951-402-3
; Sequence 3, Application US/09951402
; Patent No. US20020168752A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/951,402
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 09/306,998
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 60/084,740
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-951-402-3

```

```

Alignment Scores:
Pred. No.: 1,79e-11 Length: 2037
Score: 263.00 Matches: 146
Percent Similarity: 36.47% Conservative: 75
Best Local Similarity: 24.09% Mismatches: 178
Query Match: 6.59% Indels: 207
DB: 9 Gaps: 26

```

US-09-502-945-5 (1-2162) x US-09-951-402-3 (1-2037)

QY 11. TCGCGGTGCGGCTCTTCCAGCTCTGCGACCGCGGACCCGAA-----GGAAAG 61

```

Db 1552 SerGlnSerLeuMetValProGlnSerGlySerProGluProGluSerIleArgAsnThr 1571
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 GGTCTGCAACACGACGACGCTGCGCCAGCCATGAGCAAAAGTGGCCGAGAAAT 121
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1572 SerArgSerSerThrProAlaIlePheAla-Ser---AspProAlaThrCysProIleIl 1590
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 122 TCCGCATCAAGTGATTTCTGATTGAAATGATGC----- 158
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1590 eProgly-----CysGluThrThrIleGluIleSerLys 1601
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 159 ---AGAGAGGACTATCTCTATGATGCTGCTG----- 186
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1601 sGlyArgThrGlyLeuGlyLeu-SerIleValGlyGlySerAspThrLeuLeuGlyAlaI 1621
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 187 -----CGAATGTACCACCAG-----ACCATGAGCGTGGCGTCTCGTGG 226
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1621 leIleIleHisLeuValTyrGluGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaG 1641
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 227 GAGACCTGAAGCTGCTCATCAAT----- 249
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1641 LysAspGluIleLeuGluValAsnGlyIleAspLeuArgLysAlaThrHisAspGluAlaI 1661
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 250 -----GAACCCAGCGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 292
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1661 leAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyrArgAspGluA 1680
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 293 TCCCACTGAAGCACCAGCTGGAATATGATCAGCTGACCCCGCGCTCCAGGAAGTGA 352
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1680 laProTyrLysGluGluValCysAspThrLeuThr----- 1692
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 353 AGGAGGTGCTGACCGCTGTCACCCCGAGGCTGCGCTGAGTGTGCTGCTGCTGCTG 412
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1693 -----IleGluLeuGlnLysLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA 1711
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 413 TGGAGTTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAAAGGCGGTGAGCAGACGC 472
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1711 rGAsn---AspThrGlyValPheValSerAspIleValLysGlyIleAlaAspAlaA 1730
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 473 TCGGG---CTCCAGTGGGAGGACGATCGTCCGATCAATGATATTCATCTCCTCT 529
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1730 spGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGlyGluAspValArgAsnA 1750
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 530 GTACCATGAGGAGGTGATCAACCTCATTCGA---ACCAAGAAACTGTGTCCATCAAG 586
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1750 laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGlyThrValThrLeuGluV 1770
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 587 TGAGACATCGGCTGATCCCGTGAAGAGCTCT-----CCTGATGAG----- 630
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1770 alGlyArgIleLysAlaGlyProPheHisSerGluArgArgProSerGlnSerSerGlnV 1790
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 631 -----CCCTCACCTGGCAGTATGCTGATGATCAGT 658
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1790 alSerGluGlySerLeuSerSerPheThrPheProLeuSerGlySerSerThrSerGluS 1810
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 659 TTGTGTCGGAATCTGGGGCGTGGAGGACGCTGGGCTCC-----CCTGGAATCGGG 712
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1810 erLeu---GluSerSerSerLysLysAsnAlaLeuAlaSerGluIleGlnGlyLeuArgT 1829
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 713 AAAACAGGAGAGAAAGCTCTTCATCAGCTGGTAGGCTCCCGAGGCTTGGCTGCAGCA 772
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1829 hrValGluMetLysLys-----GlyProThrAspSerLeuGlyIleSerI 1844
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 773 TTTCAGCGGC-----CCCATCCAGAGCGCTGGCATCTTTATCAGCCATGTGAAC 823
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1844 leAlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetHetiSP 1864
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 824 CTGGCTCCCTGCTGCTGAGGTG---GGATTGGAGATAGGGGACAGAGTGTTCGAAGTCA 880
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1864 roThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleC 1884
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 881 ATGGGTCGACTTCTCTAACCTGGATCACAAGGAGGCTGTAATGTGCTGAAAATAGCC 940
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


Db	1884	ysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValAsnLeuLeuLysAsnAla-	1903
QY	941	GCAGCTGACCATCTCCATTGTAGCTGCAGCTGCCGGGAGCTGTTCATGACAGACGGG	1000
Db	1904	--SerGlySerIleGluMetGlnValValAlaGly-GlyAspValSerValValThrGly	1922
QY	1001	AGCGGCTGCAGAGCGCGGCGAGCTGAGTGCAGCGGCGAGGAGTCTTCATCGAAGAC	1060
Db	1922	-----	1922
QY	1061	GGCTGGCGATGGATCCACACAGATCCTCCAGGAGCAGCAGGAGATCGGCGGCAAGGA	1120
Db	1922	-----	1922
QY	1121	GAAGAAGAAATTGCCAGRAAGCGCAGAGAGAAATGAGAGATACCGGAAGAGATGGAAC	1180
Db	1922	-----	1922
QY	1181	AGATTGTAGAGGAGGAAGAAAGTTTAAGAAGCAATGGAAGAAGATGGGGGTCAAAGG	1240
Db	1922	-----	1922
QY	1241	AACAGCTACTCTTGCCATAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGC	1300
Db	1923	-----HisGlnGlnPuProAlaSerSer	1930
QY	1301	CAAGTGATTTCGGAAATATGAGGAAGGCTTTGACCCCTACTCTATGTTCAAC-	1354
Db	1931	SerLeuSerPhe-----ThrGlyLeuThrSerSerSerIlePheGlnAspAsp	1946
QY	1355	-----CCAGAGCAGATCATGGGAAGGATGTCGGCTCTACGCATCAAGAAGGAGGGA	1408
Db	1947	LeuGlyProGlnCys-----LysSerIleThrLeuGluArg-----GlyProAsp	1962
QY	1409	TCCTTAGACCTGGCCCTGGAGCGGCTGGACTCCCCCATTTGG------AAGTGGTC	1462
Db	1963	GlyLeuGlyPheSerIleValGlyGlyTyrglySerProHisGlyAspLeuProIleTy	1982
QY	1463	GTTTCTGCTGTATGAGCGGGGAGCTGCTGAGCGGCATGGTGCAATTGTGAAGGGGAC	1522
Db	1983	ValIysThrValPheAlaIysGlyAlaAlaSerGluAspGlyArgLeuLysArgGlyAsp	2002
QY	1523	GAGATCATGGCAATCAACGGCAAGATTGTGCAGACTACACCCCTGGCTAGGCTGACGT	1582
Db	2003	GlnIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaValAla	2022
QY	1583	GCCTGCAGAG	1594
Db	2023	IleLeuLysArg	2026

```

RESULT 9
US-09-951-401-3
; Sequence 3, Application US/09951401
; Patent No. US20020115104A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/951,401
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 09/306,998
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 60/084,740
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2037
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-951-401-3

```

Alignment Scores:					
Pred. No.:	1.79e-11	Length:	2037		
Score:	263.00	Matches:	146		
Percent Similarity:	36.47%	Conservative:	75		
Best Local Similarity:	24.09%	Mismatches:	178		
Query Match:	6.59%	Indels:	207		
DB:	10	Gaps:	26		
US-09-502-945-5 (1-2162) x US-09-951-401-3 (1-2037)					
Qy	11	TCGGGFTCGGGCTCTTCCAGCTCTGGCAGCGGCACCCGAA-----GGAACG	61		
Db	1552	SerGlnSerLeuMetValProGlnSerGlySerProGluProGluSerIleArgAsnThr	1571		
Qy	62	GGTCGTGCAACAGCAGCAGCTGGACCTGGCCCCAGCATGACCAGAAAAGTGCCCGCAGAAT	121		
Db	1572	SerArgSerSerThrProAlaIlePheAla-Ser---AspProAlaThrCysProIleIl	1590		
Qy	122	TCGGGCATAAGTGGATTTCGTGATTGANAATGATGC-----	158		
Db	1590	eProGly-----CysGluThrIleGluIleSerLy	1601		
Qy	159	----AGAGAAAGGACTATCTCTATGATGTGCTG-----	186		
Db	1601	sGlyArgThrGlyLeuGlyLeu-SerIleValGlySerAspThrLeuLeuClyAlaI	1621		
Qy	187	-----CGAATGTACCACAG-----ACCATGGACGTGCCGTGCTCTCTGG	226		
Db	1621	IleIleIleHisGluValTyrgluGluGlyAlaAlaCysLysAspGlyLeuTrpAlaG	1641		
Qy	227	GAGACCTGAAGCTGGTCATCAAT-----	249		
Db	1641	LysAspGlnIleLeuGluValAsnGlyIleAspLeuArgLysAlaThrHisaspGluAlaI	1661		
Qy	250	-----GAAACCCAGCGCTCCTGCCCTCTGTTTATGCCATTTCGCCCGCTGA	292		
Db	1661	LeAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyrrArgAspGluA	1680		
Qy	293	TCCCACCTGAAGCACACAGGTGGAAATGATACGTACGTACGCCGCCCGGGCTCCAGGAAGCTGA	352		
Db	1680	LaProTyrltyrGluGluGluValCysAspThrLeuThr-----	1692		
Qy	353	AGGAGGTGCGCTTGGACCGCTCTGCACCCCGAGGCGCTGGCTGAGTGGCTGGTGGCC	412		
Db	1693	-----IleGluLeuGlnLysLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA	1711		
Qy	413	TGGAGTTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAAGCGCGTCAGGCACACACGC	472		
Db	1711	rgAsn---AspThrGlyValPheValSerAspIleValysGlyIleAlaaspAlaa	1730		
Qy	473	TCGGG---CTCCAGGTAGGGAGCAGATCGTCCGGATCAATGGATTAATTCATCTCCCTCT	529		
Db	1730	spGlyArgLeuMetGlnGlyAspGlnIleLeuMetValasnGlyGluAspValArgAsnA	1750		
Qy	530	GTACCCATGAGGAGGTATCAACCTCATTCGA---ACCAGAAAACTGTGTCATCAAG	586		
Db	1750	laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuglyThrValThrLeuGluV	1770		
Qy	587	TGACACATCGGCCCTGATCCCGTCGAAAGCTCT-----CCTGATCAG-----	630		
Db	1770	aLGlyArgIleLysAlaGlyProPheHisSerGluArgargProSerGlnSerSerGlnV	1790		
Qy	631	-----CCCTCATCTTGGCAGTATGGATCAGT	658		
Db	1790	alserGluGlySerLeuSerSerPheThrPheProLeuSerSerGlySerSerThrSerGluS	1810		
Qy	659	TTGTCTCGGAATCTGGGGCGTCGAGCAGCGCTCTGGGTCC-----CCTGAAATCGGG	712		
Db	1810	erLeu---GluSerSerLysLysAsnAlaLeuAlaSerGluIleGlnGlyLeuArgT	1829		
Qy	713	AAACACAGGAGAGAAGGTCTTCATCAGCGCTGGTAGGCTCCGAGGCCCTTGGCTGCACGA	772		
Db	1829	hrValGluMeLysLys-----GlyProThrAspSerLeuGlyIleSerI	1844		

QY	773	TTTCCAGCGC-----CCCATCCAGAACCTGGCATCTTTATCATGCCATCTGGAAC	823
Db	1844	leAlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetMethIsp	1864
QY	824	CTGGCTCCCTGTCTGCTGAGGTG---GGATTGGAGATAGGGGACCATTTGTCGAAGTCA	880
Db	1864	roThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrIleC	1884
QY	881	ATGSGCTGCACTCTCTAACCCTGGATCACAAGGAGGCTGTAATGTGCTGAAAAATAGCC	940
Db	1884	ysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValAsnLeuLeuLysAsnAla-	1903
QY	941	GCAGCTGCACATCTCAATTGTAGCTGCGAGTCGGCGGAGCTGTTTCATGACAGACCGGG	1000
Db	1904	--SerGlySerIleGluMetGlnValValAlaGly-GlyAspValSerValValThrGly	1922
QY	1001	AGCGGCTGGCAGAGCGCGCGACGCTGAGCTGCAGCGCAGGAGCTCTCATGCAAGAAGC	1060
Db	1922	-----	1922
QY	1061	GGCTGGCGATGGATCCCAACAGATCTCCAGGACGACGAGAGATGGAGCGCAAGGA	1120
Db	1922	-----	1922
QY	1121	GAAGAAGAAATTGCCAGAGAGCAGCAGAGGAAATGAGAGATACCGGAAGGAGATGGAAC	1180
Db	1922	-----	1922
QY	1181	AGATTGTAGAGGAGGAAGAAGATTTAAGAAGCAATGGGAAGAAGACTGGGGCTCAAAGG	1240
Db	1922	-----	1922
QY	1241	AACAGTACTCTTGCTTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGC	1300
Db	1923	-----HisGlnGlnGluProAlaSerSer	1930
QY	1301	CAAAGTGATTCCGGAAATATCAGGAAGGCTTGACCCCTACTCTATGTTTCAACC-----	1354
Db	1931	SerLeuSerPhe-----ThrGlyLeuThrSerSerSerIlePheGlnAspAsp	1946
QY	1355	-----CCAGACAGATCATGGGAAGGATGTCGGCTCTCTACGCATCAAGAAGGAGGA	1408
Db	1947	LeuGlyProGlnCys-----LysSerIleThrLeuGluArg-----GlyProAsp	1962
QY	1409	TCCTTAGACTGGCCCTGGAAGCGGTGTGGACTCCCCCATTTGGG-----AAGGTGTC	1462
Db	1963	GlyLeuGlyPheSerIleValGlyGlyTyrglySerProHisGlyAspLeuProIleTyr	1982
QY	1463	GTTCCTGCTGTATCAGCGGGAGCTCTGAGCGGATGTCGTCATTTGAAAGGGGAC	1522
Db	1983	ValLysThrValPheAlaLysGlyAlaAlaSerGluAspGlyArgLeuLysArgGlyAsp	2002
QY	1523	GAGATCATGGCAATCAACGGCAAGATTGTACAGACTACACCCCTGGCTGAGGCTGACGCT	1582
Db	2003	GlnIleAlaValAsnGlyClnSerLeuGluGlyValThrHisGluGluAlaValAla	2022
QY	1583	GCCTTCAGAAAG	1594
Db	2023	IleLeuLysArg	2026

```

RESULT 10
US-09-922-101-3
; Sequence 3, Application US/09922101
; Patent No. US20020146711A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: MMS2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMS2
; CURRENT APPLICATION NUMBER: US/09/922,101
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/306,998

```

[illegible]

Db 1160 SerSerThrProArgValIleProAsnValHisAsnLysAlaAsnLysIleThr---Ser 1177
QY 646 TATGTGATCAGATTGTGTGCGAATCTGGGGCGTCCGAGGAGCGCTGGCTCCCTC--- 702
Db 1179 AsnGlnAsnGlnAspThrGlnGluLysLysGluLysArgGlnGlyThrAlaProPro 1198
QY 703 -----GGAAATCGGAAACAAAGCAAG 726
Db 1199 MetLysLeuProProTyrLysAlaLeuThrAspAspSerAspGluAsnGluGlu 1218
QY 727 AAGTCTTC-----ATCAGCTGGTAGGC----- 750
Db 1219 AspAlaPheThrAspGlnLysIleArgGlnArgTyrAlaAspLeuProGlnGluLeuHis 1238
QY 751 -----TCCGAGGCCTGGCTGCAGCATTTCCAGCGGCCCATC 789
Db 1239 IleIleGluLeuGluLysAspLysAsnGlyLeuGlyLeuSerLeuAlaGlyAsnLysAsp 1258
QY 790 CAG---AAGCCTGGCATCTTTATCACCCTGGAACCT---GGCTCCCTCTGCTGTGAG 843
Db 1259 ArgSerArgMetSerIlePheValValGlyIleAsnProGluGlyProAlaAlaAsp 1278
QY 844 GTGGGATTGAGTAGGGACACAGATTGCGAAGTCAATGGCTGCAGCTTCTCTAACTG 903
Db 1279 GlyArgMethIleGlyAspGluLeuLeuGluIleAsnAsnGlnIleLeuTyrGlyArg 1298
QY 904 GATCACAGGAGGCTGTAATGTGCTGTAATAATGCCAGCCTGACCATCTCCATTGTA 963
Db 1299 SerHisGlnAsnAlaSerAlaIleIleLysThrAlaProSer----- 1312
QY 964 GCTGACGCTGGCGGAGCTGTTTCATGACAGCGGGAGCGGTGCGCAGGCGCGCGAG 1023
Db 1313 -----LysValLysLeuValPheIleArgAsn 1321
QY 1024 CGTGAGCTGCAGCGGAGGAGCTTCTCATGCAGAGCGGTGGCGATGAGTCCAAACAG 1083
Db 1322 GluAspAlaValAsnGlnMetalValThrProPheProValProSerSerProSer 1341
QY 1084 ATCTCCAGGAGCAGCAG-----GAGTGGAGCGGCAAGAGGAGAAAAAGAA 1128
Db 1342 SerIleGluAspGlnSerGlyThrGluProIleSerSerGluLysAspGlySerLeuGlu 1361
QY 1129 ATTGCCAAGAACGACGACAGAGAAATGAGATACCGAGAGGAGATGGAACAGATTGTA 1188
Db 1362 ValGlyIleLysGlnLeuProGluSerGluSerPheLysLeuAlaValSerGlnMet--- 1380
QY 1189 GAGGAGGAAGAAGTTTAAGAACGAATGGGAAGAGACTGGGCTCAAAGAACAGCTA 1248
Db 1381 ---LysGlnGlnLysTyr----- 1385
QY 1249 CTCTTGCTAAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTGA 1308
Db 1386 -----ProThrLysValSerPheSerSerGlnGluIleProLeuAlaProAlaSer 1403
QY 1309 TTTCCGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTC----- 1351
Db 1403 rTyrHisSerThrAspAlaAspPheThrGlyTyrGlyGlyPheGlnAlaProLeuSerVa 1423
QY 1352 -----ACCCAGACGATCATGGGAAGATGTCGGGTCTCATGCATCAAGAA 1401
Db 1423 LaspProAlaThrCysProIleValProGlyGlnGluMetIleIleGluIleSerLysG 1443
QY 1402 GGAGGGATCTTAGACTCGCCCTGGAAGCGGTGTGAGCTCCCCCATTTGGGAAGTGTGT 1461
Db 1443 YArgSerGlyLeuGlyLeuSerIleValGlyLysAspThrProLeuAsnAlaIleVa 1463
QY 1462 CGTTTCTGCTGTATGACCGGGAGCTCTCAGCGGCATGGTGGCATTTGTGAAGGGGA 1521
Db 1463 ILeHisGluValTyrGluGluGlyAlaAlaAlaArgAspGlyArgLeuTrpAlaGlyAs 1483
QY 1522 CGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTGGCTGAGCTGACGC 1581
Db 1483 pGlnIleLeuGluValAsnGlyValAspLeuArgAsnSerHisGluGluAlaIleTh 1503

Qy	1582	TGCCCTGCAGAGCCTCGAATACAGGCGGGGACTGGATGACCTTGTGGTTCGCCTG	1641
Db	1503	raLaLeuArg- :::--GlnThrProGlnLysValArgLeuValVal	1516
Qy	1642	CCCCCAAGAGGATGACGATGAGCTGACCTTCCTTG- :::--TyrArgAspGluAlaHisTyrArgAspGluGluAsnLeuGluLeuPhe	1678
Db	1517	-----CTGAATGCCAAAGGGGA- :::	1532
Qy	1679	eProValAspLeuGlnLysLysAlaGlyArgGlyLeuGlyLeuSerIleValGlyLysAr	1552
Db	1532	-----AACCAATTCACGCTTACGAAACAGTGGAGCTCCGCC	1734
Qy	1697	gAsnGlySerGlyValPheIleSerAspIleValLysGlyGlyAlaAlaAspLeuAspG	1572
Db	1552	gAsnGlySerGlyValPheIleSerAspIleValLysGlyGlyAlaAlaAspLeuAspG	1572
Qy	1735	CCACTCTGTGAACACAAAGCCTCGGACCGCTTGAGAGAGCCACATGACACACACAG	1794
Db	1572	YArgLeuIleGlnGlyAspGlnIleLeuSerValAsnGlyGluAspMetArg- ::: :::--	1589
Qy	1795	ATGCGATCCTTTGGGACCTCAATCTATCACCCAGGAATCTCAAACTCCTTTGGCCCTGAA	1854
Db	1590	-----AsnAlaSerGlnGluThrValAlaThrIleLeuLysCysAl	1603
Qy	1855	CCAGGCCGAGATGAAGACAGCTCGGGCCACTTTTGTGAAGCCCAATGTGGAGGAAGGG	1914
Db	1603	agInGlyLeuValGlnLeuGluIleGlyArgLeuArgAlaGlySer--TriPThr- :::	1620
Qy	1915	AGCAGCGACGCGTTTGGGAGAAGATCTCAAGATCCAGACTCTCATCTCTTCTCTCTGGC	1974
Db	1621	SerAlaArgThrThrSerGlnAsnSerGlnGlySerGlnSerAlaHisSerSer- :::	1974
Qy	1975	CCAGTGAATTTGGTCTCTCCCAAGCTTTTGGGGACTCCTT	1639
Db	1640	-----CysHisProSerPheAlaProValIle	1648

```

RESULT 12
US-09-997-977-3
; Sequence 3, Application US/09997977
; Publication No. US2003002728A1
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: MMSC1 - An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC1 Gene
; CURRENT APPLICATION NUMBER: US/09/997,977
; CURRENT FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 09/233,086
; PRIOR FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: US 60/071,861
; PRIOR FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 3
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-997-977-3

```

Alignment Scores:	
Pred. No.:	1.53e-10
Score:	250.50
Percent Similarity:	36.03%
Best Local Similarity:	20.53%
Query Match:	6.28%
DB:	9
Length:	1881
Matches:	155
Conservative:	117
Mismatches:	281
Indels:	202
Gaps:	28

US-09-502-945-5 (1-2162) x US-09-997-977-3 (1-1881)

Qy 70 AACGACGCAGCTGGACCTGGCCCGCCAGCCATGGACCGAAAGTGCCCGGAGAAATTCGGCAT 129

Db 991 AsnAspValGlnGlyProSerLeuLeuIleAspLeuProValValAlaGlnArgArgGlu 1010
Qy 130 AAGGTGGATTCTCTGATTGAAATATCATCAGAGAGGACTACTCTATGATGTGCTCGGA 189
Db 1011 GlnGluAsp-----LeuPro 1015
Qy 190 APTGATC---CACAGACCATGACCTGCCCTGCTCGTGGGAGACCTGAAGCTGTGTCATC 246
Db 1016 LeuTyrGlnHisGlnAla-----ThrArgValIle 1025
Qy 247 AATGAACCCAGCGCTCTGCTGCTGTTGATGCCATTCGCCCTGATCCCACTGAAGCAC 306
Db 1026 SerLysAlaSer-----AlaTyrThrGlyMetLeuSerSerArgTyr 1039
Qy 307 CAGGTGGATATGATCAGCTGACCGCCCGCGCTCCAGGAAGCTGAAGGAG----- 357
Db 1040 AlaThrAspThrCysGlnLeuProGluArgGluGluGlyGluGluThrProAsn 1059
Qy 358 -----GTGCGTCTGGACCGCTCTGCACCCCGAAGCGCTC 390
Db 1060 PheSerHisTyrGlyProProArgIleValGluIlePheArgGluProAsnValSerLeu 1079
Qy 391 GCCTGAGTGTGGTGGG-----CTGGAGTTGGC----- 423
Db 1080 GlyIleSerIleValGlyGlnThrValIleLysArgLeuLysAsnGlyGluGluLeu 1099
Qy 424 TGTGGCTCTTCATCTCCACCTCATCAAGCGGTGCAGGCAGCAGCGTC---GGGCTC 480
Db 1100 LysGlyIlePheIleLysGlnValLeuGluAspSerProAlaGlyLysThrAsnAlaLeu 1119
Qy 481 CAGGTAGGGAGCAGATCGTCGGGATCAATGGATATTCATCTCTCTGTACCCATGAG 540
Db 1120 LysThrGlyAspLysIleLeuGluValSerGlyValAspLeuGluAsnAlaSerHisser 1139
Qy 541 GAGGTCACTCAACCTCATTCGACCCAG---AAACTGTGTCTCCATCAAGTGAAGACATC 597
Db 1140 GluAlaValGluAlaIleLysAsnAlaGlyAsnProValIleValGlnSerLeu 1159
Qy 598 GGC-----CTGATCCCGTGAAGACTCCTCGATGAGCGCCCTCACTTGGCAG 645
Db 1160 SerSerThrProArgValIleProAsnValHisAsnLysAlaAsnLysIleThr---Ser 1178
Qy 646 TATGTGATCAGTTGTGCGGAATCTGGGGCGTCCGAGCAGCGCTGGGCTCCCTC--- 702
Db 1179 AsnGlnAsnGlnAspThrGlnGluLysLysGluLysArgGlnGlyThrAlaProPro 1198
Qy 703 -----GGAAATCGGAAACAAAGGAGAG 726
Db 1199 MetLysLeuProProTyrLysAlaLeuThrAspAspSerAspGluAsnGluGlu 1218
Qy 727 AAGGTCTTC-----ATCAGCTGTGTAGGC----- 750
Db 1219 AspAlaPheThrAspGlnLysIleArgGlnArgTyrAlaAspLeuProGlyGluLeuHis 1238
Qy 751 -----TCCCGAGGCTTGGCTGCGACATTTCCAGCGCCCGCCATC 789
Db 1239 IleIleGluLeuGluLysAspLysAsnGlyLeuGlySerLeuAlaGlyAsnLysAsp 1258
Qy 790 CAG---AAGCCTGGCATCTTATCAGCCATGTGAACCT---GGCTCCCTGTCTGTGAG 843
Db 1259 ArgSerArgMetSerIlePheValValGlyIleAsnProGluGlyProAlaAlaAsp 1278
Qy 844 GTGGGATTGAGATAGGGACCAAGATTGTCGAAGTCAATGGCGCTCTCTCAACCTG 903
Db 1279 GlyArgMetHisIleGlyAspGluLeuLeuGluIleAsnAsnGlnIleLeuTyrGlyArg 1298
Qy 904 GATCACAAGAGGCTGTAAATGTGCGAATAATAGCCGAGCGCTGACCATCTCCATTGTA 963
Db 1299 SerHisGlnAsnAlaSerAlaIleIleLysThrAlaProSer----- 1312
Qy 964 GCTGCAGCTGGCGGGAGCTGTTTCATCAGACAGCCGGAGCGGTGGCAGAGCGCGCAG 1023
Db 1313 -----LysValLysLeuValPheIleArgAsn 1321

Qy 1024 CGTGAGCTGCAGCGGCGAGGCTTCTCATCAGAAAGCGGTGGCGATGGAGTCCACAAG 1083
Db 1322 GluAspAlaValAsnGlnMetAlaValThrProPheProValProSerSerProSer 1341
Qy 1084 ATCTCCAGGAGCAGCAG-----CAGATGGAGCGCAAGAGAAAGAA 1128
Db 1342 SerIleGluAspGlnSerGlyThrGluProIleSerSerGluGluAspGlySerLeuGlu 1361
Qy 1129 ATTCCCAAGAGCAGCAGAGAGAAATCAGAGATACCGGAAGGAGATGGAAACGATTGTA 1188
Db 1362 ValGlyIleLysGlnLeuProGluSerGluSerPheLysLeuAlaValSerGlnMet--- 1380
Qy 1189 GAGGAGGAAGAGAGTTTAAGAAGCAATGGGAAGACAGCTGGGCTCAAAGAAACAGCTA 1248
Db 1381 ---LysGlnGlnLysTyr----- 1385
Qy 1249 CTCCTTGCCTAAACCCATCAGTCTGAGGTACACCCAGTACCCCTTCGCAACCAAGTGA 1308
Db 1386 ---ProThrLysValSerPheSerSerGlnGluIleProLeuAlaProAlaSerSe 1403
Qy 1309 TTTCGGGAAATATGAGGAAGCGTTTGACCCCTACTCTATGTC----- 1351
Db 1403 rTyrHisSerThrAspAlaAspPheThrGlyTyrGlyGlyPheGlnAlaProLeuSerVa 1423
Qy 1352 -----ACCCAGAGCAGATCATGGGAAGGATGTCGCGCTCCTACCATCAAGAA 1401
Db 1423 LasProAlaThrCysProIleValProGlyGlnGluMetIleIleGluIleSerLysG 1443
Qy 1402 GGAGGATCCTTAGACCTGGCCCTGGAAGCGGTGTGAGCTCCCCATTTGGGAAGGTGGT 1461
Db 1443 yArgSerGlyLeuGlyLeuSerIleValGlyLysAspThrProLeuAsnAlaIleVa 1463
Qy 1462 CGTTTCTCTGTATGAGCGGAGCTGTGAGCGGATGGTGGCATTTGTGAAGGGGA 1521
Db 1463 IleHisGluValTyrGluGluGlyAlaAlaAlaArgAspGlyArgLeuTyrAlaGlyAs 1483
Qy 1522 CGAGATCATGCAATCAACGGCAAGATTGTGACAGACTACACCTCGCTGAGGTGAGCGC 1581
Db 1483 pGlnIleLeuGluValAsnGlyValAspLeuArgAsnSerSerHisGluGluAlaIleTh 1503
Qy 1582 TGCCCTGAGAGCGCTGGAATCAGGGCGGAGCTGGATCGACCTGTGGTGGTGGCTG 1641
Db 1503 rAlaLeuArg-----GlnThrProGlnLysValArgLeuValVal----- 1516
Qy 1642 CCCCCAAAGAGTATGAGTACGATGAGCTGACCTCTCTG----- 1678
Db 1517 -----TyrArgAspGluAlaHisTyrArgAspGluGluAsnLeuGluIlePh 1532
Qy 1679 -----CTGAAGTCCAAAGGGGA----- 1696
Db 1532 eProValAspLeuGlnLysLysAlaGlyArgGlyLeuGlyLeuSerIleValGlyLysAr 1552
Qy 1697 -----AACCAATTCACCGCTTAGGAACACAGTACGCTCGGCC 1734
Db 1552 GAsnGlySerGlyValPheIleSerAspIleValLysGlyGlyAlaAlaAspLeuAsp 1572
Qy 1735 CCACCTCGTGAACACAAAGCGCTTCGAGAGAGCGCCACATGACACACACCAG 1794
Db 1572 yArgLeuIleGlnGlyAspGlnIleLeuSerValAsnGlyGluAspMetArg----- 1589
Qy 1795 ATGGCATCCTTGGGACCTGAATCTATCACCAGGAATCTCAACCTCCCTTTGGCCCTGAA 1854
Db 1590 -----AsnAlaSerGlnGluThrValAlaThrIleLeuLysCysAl 1603
Qy 1855 CGAGGCCAGATGAAGAACAGCTCGGGCCACTTTTTTGAAGGCCAATGTGGAGGAAGG 1914
Db 1603 aGlnGlyLeuValGlnLeuGluIleGlyArgLeuArgAlaGlySer---TrpThr----- 1620
Qy 1915 AGCAGCCAGCCGCTTTGGGAGAGATCTCAAGGATCTCAGAGTCTCATCTCTCTCTGCGC 1974
Db 1621 SerAlaArgThrThrSerGlnAsnSerGlnGlySerGlnGlnSerAlaHisSer--- 1639

QY 1975 CCAGTGAATTTGGTCTCTCCAGCTTTGGGGACTCCTT 2013
Db 1640 -----CysHisProSerPheAlaProValIle 1648
|||||

RESULT 13

US-09-864-761-39112
; Sequence 39112, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006684
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39112
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005137.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: SWISSPROT HIT: P31007, EVALUE 1.10e-01
; OTHER INFORMATION: EST_HUMAN HIT: BE731162.1, EVALUE 2.40e-02
US-09-864-761-39112

Alignment Scores:
Pred. No.: 3.39e-10 Length: 46
Score: 237.00 Matches: 46
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.94% Indels: 0
DB: 10 Gaps: 0

US-09-502-945-5 (1-2162) x US-09-864-761-39112 (1-46)

QY 346 AAGCTGAAGGAGGTGCTGACCGTCTGCACCCGAGCGCTCGGCTGAGTGTGCGT 405
Db 1 LysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeuSerValArg 20
QY 406 GTGCGCTGGAGTTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAAAGGGGTGAGCA 465
Db 21 GlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuLeuLysGlyGlyGlnAla 40
QY 466 GACACGCTCGGCTCCAG 483
Db 41 AspSerValGlyLeuGln 46

RESULT 14

US-09-919-497-59
; Sequence 59, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59
; LENGTH: 767
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-59

Alignment Scores:
Pred. No.: 1.72e-09 Length: 767
Score: 234.50 Matches: 144
Percent Similarity: 38.10% Conservative: 104
Best Local Similarity: 22.12% Mismatches: 210
Query Match: 5.88% Indels: 193
DB: 10 Gaps: 34

US-09-502-945-5 (1-2162) x US-09-919-497-59 (1-767)

QY 340 TCCAGGAGCTGAAGGAGGTGCGTCTGCACCGTCTGCACCCGAGCGCTCGGCTGAGT 399
Db 198 AlaGluLysValMetGluIleLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 216
QY 400 GTGCGTGTGGTGC-----CTGGAGTTTGGCTGTGGCTCTTCATCTCCAC 444
Db 217 IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 236
QY 445 CTCATCAAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 501
Db 237 IleIleGluGlyGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 256
QY 502 CGGATCATGATATTCATCTCTCTCTGTACCCATGAGGAGGAGGAGGAGGAGGAGGAG 561
Db 257 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaAlaLeuLys 276
QY 562 ---ACCAAGAAACTGTCTCCATCAAGTG-----AGACACATCGGCTGATCCCC 609
Db 277 AsnThrTyrAspValValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 296

Qy	610	GTGAAAGACTCTCCTGATGAGCCCTCACTGG---CAGTATGTGGATCAGTTTGTCGTGC	666
Db	297	SerTyAlaProProAspPheThrThrSerTySerGlnHisLeuAsnGluIleSer	316
Qy	667	GAATCGGGGGCGTCGACGCCAGCTCGGC---	696
Db	317	HISser-----SerTyLeuGlyThrAspTyProThrAlaMetThrProThr	332
Qy	697	-----TCCCTCGAANTCGGAA-----AACAAg	720
Db	333	SerProArgTySerProValAlaLysAspLeuLeuGlyGluGluAspIleProArg	352
Qy	721	GAGAAGAAGTCTTCATCAGCTGTGTAGGTCGCCAGGCGCTGCTCCACATTTCCAGC	780
Db	353	GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly	372
Qy	781	GGCCCCATCCAGAAGCTCGCATCTTTATCAGCCATG-----AAACCTGGCTCC	831
Db	373	GlyGlu---AspGlyGluGlyPheIleSerPheIleLeuAlaGlyGlyProAlaAsp	391
Qy	832	CTGTCTGCTGAGTGGGATTGGAGATAGGGACCAGATTGTCGAAGTCAATGGCGTCGAC	891
Db	392	LeuSerGlyGlu-----LeuArgIysGlyAspGlnIleLeuSerValasnGlyValasp	409
Qy	892	TTCTTAACCTGGATCACAGGAGGCTGTAATGTGCTGAAAAATACCCGACCGCTGACC	951
Db	410	LeuArgAsnAlaSerHisGluGlnAlaLleAlaLeuLysAsnAla---GlyGlnThr	428
Qy	952	ATCTCCATTCTAGCTCGACTGCGCGGAGCTGTTTCATGACAGACCGGAGCGGCTGGCA	1011
Db	429	ValThrIleIlealagln-----TyrLysProGluGluTySerArgPhe	443
Qy	1012	GAGCGCGGCGAGCTGAGCTGCAGCGGACGAGGCTTCTCATCGACAAGCGGCTGGCGATG	1071
Db	444	GluAlatylsileHisAspLeuArgGluGln----LeuMetAsnSerSerLeuGlySer	461
Qy	1072	GAGTCCAACAGATCTCTCCAGNGACGACGAG-----ATGCAG	1110
Db	462	GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrlleArgAlaLeuPheAsp	481
Qy	1111	CGGCAAGGAGAAAGAAATTCGCCAGAAGCGACGAGGAGAAATGAGAGATACCGGAAG	1170
Db	482	TyrAspLysThrLysAspcysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp	501
Qy	1171	GAGTAGAACAGATTCTGAGGAGGAGGAAGAGATTT-----	1206
Db	502	ValLeuHisValIleAspAlaSerAspGluGluTrpTrpGlnAlaArgArgValHisSer	521
Qy	1207	-----AAGAAGCAA	1215
Db	522	AspSerGluThrAspAspIleGlyPheIleProSerLysArgArgValIGluArgArgGlu	541
Qy	1216	TGGCAA-----GAAGACTCGGCTCA-----AAGGAA	1242
Db	542	TrpSerArgLeuLysAlaLysAspTrpGlySerSerGlySerGlnGlyArgGluAsp	561
Qy	1243	CAGCTACTCTTGCTAAAACCATCATCT---GCTGAGGTACACCCAGTACCCCTTCGCAAG	1299
Db	562	SerValLeuSerTyrgluThrValThrGlnMetGluValHisTyrAlaArg-----	578
Qy	1300	CAAAGTGATTTCGGNAATATAGGAGAGGCTTTGACCCCTACTCTATGTTCACCCCAGA	1359
Db	579	-----ProIleIleLeuGlyProThr	586
Qy	1360	CGCATCATGGGNAGGATGTCGGGCTCTACGCATCAAGAGGAGGATCCTTAGACCT	1419
Db	587	LysAspargala-----AsnAspAspLeuLeuSerGlu	597
Qy	1420	GGCCCTGGAGGGCGGTGTGGACTCCCCCATTTGGGAAGGTGGTCTTCTGCTGTGTATGA	1479
Db	598	PheProaspyls-----	601
Qy	1480	CGGGGAGCTCTGAGCGGCGATGTTGGCATTTGTGAAAGG---GGACGAGATCATGGCAAT	1536

Search completed: March 21, 2003, 13:58:21
Job time : 52.1498 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:32:38 ; Search time 38.5119 Seconds
(without alignments)
10793.685 Million cell updates/sec

Title: US-09-502-945-5
Perfect score: 3989
Sequence: 1 cctggccggtcggtgcg.....tattttccagcttaaaaaa 2162

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlp
-O/cgn2_1/USPRO_pool/US09502945/runat_14032003_101059_19127/app_query.fasta_1.10979
-DB=PIR_73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 -CGN_1_1_263 -runat_14032003_101059_19127 -NCPU=6 -ICPU=3
-NO_XLPYX -NO_MAP -LARGESQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	269.5	6.8	2054	2 T46612	multi PDZ domain p
2	259.5	6.5	2055	2 T30259	multiple PDZ domai
3	244.5	6.1	720	2 A45436	synapse-associated
4	243	6.1	1012	2 T23160	hypothetical prote
5	236.5	5.9	926	2 I38756	homolog of Drosoph
6	236	5.9	1745	2 A46431	tight junction-ass
7	234.5	5.9	767	2 T09599	postsynaptic densi
8	232.5	5.8	724	2 JH0800	postsynaptic densi
9	232	5.8	904	2 I38757	homolog of Drosoph
10	227	5.7	2450	2 S71625	protein-tyrosine-p
11	226	5.7	911	2 I56552	synapse-associated
12	226	5.7	1131	2 T15617	hypothetical prote
13	225	5.6	1736	2 A47747	tight junction pro
14	218.5	5.5	852	2 T10811	channel associated

15	213.5	5.4	870	2 G01974	channel associated
16	211	5.3	1112	2 T32733	AMPA glutamate rec
17	206	5.2	723	2 T14765	hypothetical prote
18	204.5	5.1	2466	2 T16729	protein-tyrosine-p
19	204	5.1	1367	2 T13703	tamA protein - fru
20	203.5	5.1	628	2 T09458	numb-binding prote
21	203.5	5.1	728	2 T09457	numb-binding prote
22	203	5.1	2294	2 T16730	protein-tyrosine-p
23	202.5	5.1	1277	2 T14152	synaptic scaffold
24	194.5	4.9	450	2 G01158	tyrosine kinase ac
25	193.5	5.0	2715	2 T13049	eyelid - fruit fly
26	192	4.8	2490	1 A54971	protein-tyrosine-p
27	191.5	4.8	1256	2 JEO209	brain-specific ang
28	190	4.8	1337	2 T13948	atypical protein k
29	189.5	4.8	2172	2 T20145	hypothetical prote
30	188.5	4.7	358	2 I46532	protein co-factor
31	188	4.7	1464	2 T13716	bazooka gene prote
32	185	4.6	1116	2 I54378	gene X104 protein
33	183	4.6	1281	2 T00346	hypothetical prote
34	181.5	4.6	1163	2 JEO366	tight junction pro
35	179	4.6	660	1 Q0BE3	BHLF1 protein - hu
36	179	4.5	1171	2 T42372	probable guanylate
37	176.5	4.5	1006	2 T42731	atrophin-1 related
38	171	4.3	817	2 T03852	protein phosphatas
39	171	4.3	1553	2 T03301	rab3 effector prot
40	167.5	4.2	1373	1 A43291	collagen alpha 2(I
41	166	4.2	960	1 A39651	discs-large tumor
42	164.5	4.1	1414	1 S23809	collagen alpha 2(I
43	164	4.1	640	2 T08179	LRG5 protein - chl
44	164	4.1	1453	2 S21626	collagen alpha 1(I
45	164	4.1	1464	1 CGH01S	collagen alpha 1(I

ALIGNMENTS

RESULT 1

T46612
multi PDZ domain protein 1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 21-Jul-2000
C:Accession: T46612
R:Ullmer, C.; Schmuck, K.; Figge, A.; Lubbert, H.
FEBS Lett. 424, 63-68, 1998
A:Title: Cloning and characterization of MUPP1, a novel PDZ domain protein.
A:Reference number: 223104; MUID:98196865; PMID:9537516
A:Accession: T46612
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2054 <ULL>
A:Cross-references: EMBL:AJ001320; NID:g2959978; PIDN:CAA04681.1; PID:g2959979
A:Experimental source: brain
C:Genetics:
A:Gene: MUPP1

Alignment Scores:	8.2e-10	Length:	2054
Pred. No.:	269.50	Matches:	120
Score:	38.20%	Conservative:	71
Percent Similarity:	24.00%	Mismatches:	161
Best Local Similarity:	6.76%	Indels:	148
Query Match:	2	Gaps:	17
DB:			

US-09-502-945-5 (1-2162) x T46612 (1-2054)

QY	157	GCAGAGAGGAGGATCTCTATGATGTCGCAATGTACCAACGACGACGACGTGGCC	216
Db	1672	AlaThrHisAspGluAlaIleAsnValLeuArgInThrProGlnArgVal	1688
QY	217	GTGCTCGTGGGAGACCTGAAGCTGGTCAATCAATGAACCCAGCCGCTCTGCTGTGTGAT	276
Db	1689	-----ArgLeuThrLeuTyr----	1693
QY	277	GCAATTCGGCGCTGATCCCACTGAAGCACCAGTGGAATATGATCAGTCGACCCCGG	336


```
Db 1694 ----ArgaspGluAlaProTyrLysGluGluAspValCysAspThrPheThr----- 1709
QY 337 CGCTCCAGGAGCTGAAGGAGCTGCTGACCGTCTGACCCCGAGGCTCGGCGCTG 396
Db 1710 -----ValGluLeuGlnLysArgProGlyLysGlyLeuGlyLeu 1722
QY 397 AGTGTGGCTGGTGGCTGGAGTTGGTGTGGGCTCTTCATCTCCACCTCATCAAGGC 456
Db 1723 SerileValGlyLysArgAsn--AspThrGlyValPheValSerAspIleValLysGly 1741
QY 457 GGTGAGGAGACAGCTGGG---CTCAGTAGGAGGAGATCGTCGGATCAATGGA 513
Db 1742 GlyIleAlaAspAlaAspGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGly 1761
QY 514 TATTCCATCTCTCTGACCCATGAGGAGTCAACCTCATTCGA--ACCAAGAAA 570
Db 1762 GluaspValArgAsnAlaThrGlnGluAlaValAlaAlaLeuLysCysSerLeuGly 1781
QY 571 ACTGTGTCATCAAGTGAGACATCGCTGATCCCGTGAAAGCTCTCTGATGAG 630
Db 1782 ThrValThrLeuGluValGlyArgIleLysAlaAlaProPheHisSerGluArgPro 1801
QY 631 CCCCTCACTGGCAGTATGTGATGAGTGTGTCGGAATCTGGGGCTCGGAGGAGC 690
Db 1802 SerGlnSerSerGlnValSerGluSerSerLeuSerPheSerLeuProArgSerGly 1821
QY 691 CTGGGCTCCCTGGAATCGGGAACAGGAGAGAAG----- 729
Db 1822 IleHisThrSerGluSerSerAlaLysLysAlaLeuAlaSerGluIle 1841
QY 730 -----GTCTTCATAGCCTGCTGAGCTCCCGAGGCTTGGCTGCAAGATT 774
Db 1842 GlnGlyLeuArgThrValGluIleLysLysGlyProAlaAspAlaLeuSerIle 1861
QY 775 TCCAGGCG-----CCATCCAGAGCTGCGCATCTTATCAGCCATGTGAACCT 825
Db 1862 AlaGlyGlyValGlySerProLeuGlyAspValProIlePheIleAlaMetHisPro 1881
QY 826 GGCTCCCTGCTGCTGAGGTG---GGATTGAGATAGGGAGCAGATGTCAAGTCAAT 882
Db 1882 AsnGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgIleValThrLys 1901
QY 883 GCGGTGAGCTTCTTAACCTGATCACAAGAGGCTGTAATGTGCTGAAATAGCCGC 942
Db 1902 GlyThrSerThrAspGlyMetThrHisThrGlnAlaValAsnLeuMetLysAsnAla--- 1920
QY 943 AGCCTGACCATCTCCATGTAGCTGACGTGCGCGGAGCTGTTTCATGACAGACCGGAG 1002
Db 1921 SerGlySerIleGluValGlnValAlaGlyAspVal----- 1934
QY 1003 CGGCTGGCAGGCGCGGCGGCTGAGCTGCGAGCGGAGGAGCTCTCATGAGAGCGG 1062
Db 1935 SerValValThrGlyHisGlnGlnGluLeuAlaAsnProCysLeuAlaPhe----- 1951
QY 1063 CTGGCGATGAGTCCCAACAGATCTCTCCAGGAGCAGCAGGAGATGGAGCGCAAGGAGA 1122
Db 1952 ThrGlyLeuThrSerSerThrIle----- 1959
QY 1123 AAGAAATTTGCCAGAGCAGCAGAGGAAATGAGAGATACCGGAGGAGATGGAACAG 1182
Db 1959 ----- 1959
QY 1183 ATTGTAGAGGAGAGAGAAATTTAAGAACCAATGGNAGAGACTGGGCTCAAGGAA 1242
Db 1960 -----PheProAspAspLeuGlyProProGln 1968
QY 1243 CAGCTACTCTGCTAAACCACTACTGCTGAGGTACACCCAGTACCCCTCGCAAGCCA 1302
Db 1969 Ser-----LysThrIleThr----- 1973
QY 1303 AGTGTATTTCCGGAATATGAGGAGGCTTTGACCCCTACTCTATGTTACCCCGAGCA 1362
```

```
Db 1974 -----Leu 1974
QY 1363 GATCATGGGAAGATGTCGGCTCCTACGCATCAAGAAGAGGATCCTTAGACCTGGC 1422
Db 1975 AspArg-GlyProAsp-----GlyLeuGlyPhe 1984
QY 1423 CTGGAAGAGCGGTGTGAGCTCCCGCATTTGG-----AAGGTGGTGTCTCTCTCTGTA 1476
Db 1984 rIleValGlyTyrGlySerProHisGlyAspLeuProIleThrValLysThrValPhe 2004
QY 1477 TGAGCGGGAGCTGCTGAGCGGCTGTCATTGTGAAGGGAGGAGATCATGGCAAT 1536
Db 2004 eAlaLysGlyAlaAlaAlaGluAspGlyArgLeuLysArgGlyAspGlnIleAlaVal 2024
QY 1537 CAACGCAAGATTTGACAGACTACACCTGGCTGGCTGAGCTGAGCTCCCTCGCAGAAG 1594
Db 2024 lAsnGlyGlnSerLeuGluGlyValThrHisGluGluAlaValAlaIleLeuLysArg 2043

RESULT 2
T30259
multiple PDZ domain protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30259
R:Simpson, E.H.; Suffolk, R.; Jackson, I.J.
Genomics 59, 102-104, 1999
A:Title: Identification, sequence, and mapping of mouse multiple PDZ domain protein 9
A:Reference number: Z20797; MUID:99326529; PMID:10395806
A:Accession: T30259
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2055 <SIM>
A:Cross-references: EMBL:AJ131869; NID:g4150877; PIDN:CAA10523.1; PID:g4150878
A:Experimental source: strain C57/BL6 x CBA Fl; whole brain
C:Genetics:
A:Gene: mpdz
A:Map position: 4

Alignment Scores:
Pred. No.: 3,82e-09 Length: 2055
Score: 259.50 Matches: 117
Percent Similarity: 37.45% Conservative: 71
Best Local Similarity: 23.31% Mismatches: 162
Query Match: 6.51% Indels: 152
DB: 2 Gaps: 19

US-09-502-945-5 (1-2162) x T30259 (1-2055)
QY 157 GCAGAGAAGGACTATCTATGATGCTGCTGCAATGTACCACAGACCATGACGTGGCC 216
Db 1673 AlaThrHisAspGluAlaIleAsnValLeuArgGlnThrProGlnArgVal----- 1689
QY 217 GTGCTGCTGGGAGACCTGAAGCTGTGCTATCAATGAACCCAGCGCTCTCTCTTTGAT 276
Db 1690 -----ArgLeuThrLeuTyr--- 1694
QY 277 GCCATTCGGCGCTGATCCCACTGAAGCACCAAGTGGATATGATCAGCTGACCCCGG 336
Db 1695 -----ArgAspGluAlaProTyrLysGluGluAspValCysAspThrPheThr----- 1710
QY 337 CGCTCCAGGAGCTGAAGGAGTGGCTGACCGCTGACCGCTGACCCCGAGGCTCGGCTG 396
Db 1711 -----IleGluLeuGlnLysArgProGlyLysGlyLeuGlyLeu 1723
QY 397 AGTGTGCTGGTGGCTGGAGTTGGCTGTGGCTCTTCTATCTCCATCTCATCAAGGC 456
Db 1724 SerIleValGlyLysArgAsn---AspThrGlyValPheValSerAspIleValLysGly 1742
QY 457 GGTGAGGAGACAGCTGGG---CTCAGGTAGGAGGAGATCGTCCGATCAATGGA 513
Db 1743 GlyIleAlaAspAlaAspGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGly 1762
QY 514 TATTCCATCTCTCTGATACCCATGAGGAGTCAACCTCATTCGA---ACCAGAAA 570
```


[illegible]

Dd 2023 eaValAsnGlyGlnSerLeuGluGlyValThrHisGluAlaValAlaIleLeuLy 2043

Qy 1591 GAAG 1594
:::
Dd 2043 sArg 2044

RESULT 3
A45436
synapse-associated protein SAP90 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000
C:Accession: A45436
R:Kistner, U.; Wenzel, B.M.; Veh, R.W.; Cases-Langhoff, C.; Garner, A.M.; Appeltauer,
J. Biol. Chem. 269, 4580-4583, 1993
A:Title: SAP90, a rat presynaptic protein related to the product of the Drosophila tu
A:Reference number: A45436; MUID:93186749; PMID:7680343
A:Accession: A45436
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-720 <KIS>
A:Experimental source: brain
A>Note: sequence extracted from NCBI backbone (NCBIP:126555)
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
F:70-148/Domain: GLGF domain homology <GLG1>
F:165-243/Domain: GLGF domain homology <GLG2>
F:435-493/Domain: SH3 homology <SH3>
F:531-708/Domain: guanylate kinase homology <GKI>

Alignment Scores:

Pred. No.:	2.61e-08	Length:	720
Score:	244.50	Matches:	144
Percent Similarity:	38.18%	Conservative:	103
Best Local Similarity:	22.26%	Mismatches:	211
Query Match:	6.13%	Indels:	189
DB:	2	Gaps:	33

US-09-502-945-5 (1-2162) x A45436 (1-720)

Qy 340 TCCAGGAAGCTGAAGGAGGTGGTCTGCAGCGTCGCCACCCGAAAGGCTCGGCGTAGT 399
::: |||||: |||||:||||: ||: |||||:|||||: |||
Dd 155 AlaGluysValMetGluIleLysLeuIlefs---GlyProlysGlyLeuGlyPheSer 173

Qy 400 GTGGGTGGTGCG-----CTGAGTTGGCTGTGGCTTTCATCTCCCAC 444
::: |||||: :::
Dd 174 IlealagGlylvalGlyAsnGlnHisfileProGlyAspnSerileTyValThrLys 193

Qy 445 CTCATCAAGCGCTCAGGCAGACAGCGTGGG---CTCCAGTAGGGGACGATCGTC 501
:::||||:|||||: ||| ||| |||||:|||||:|||||:
Dd 194 llelleGluGlylAlaAlaHisLysaspGlyArgLeuGlnIleGlyAspLysIleu 213

Qy 502 CGGATCAATGATATTCCATCTCTCTGTACCATGAGGAGTCATCAACCTCATTCGA 561
:::|||||: ||: |||||: ||: |||||: |||||:
Dd 214 AlalaasnServalGlyLeuGluaspValmethHisgluasPalalaIleLeuLys 233

Qy 562 ---ACCAAGAAAACCTGTGTCATAAGTG-----AGACACATCGGCGCTGATCCCC 609
||| ||| :||| ||||| :|||
Dd 234 AsnthrtyrAspValvalTyrlauLysvalalalysProserAsnalatyrlauSerAsp 253

Qy 610 GTGAAAGCTCTCTGTAGTACGCCCTCACITGG---CAGTATGTGGATCAGTTGTGTCG 666
::: |||||: |||||: |||||: |||||: |||||: |||||:
Dd 254 SerTyralaProAspilleThrrSerTyrserrglnHisLeuaspAsnGluIleSer 273

Qy 667 GAACTGGGGCGTGGCAGGCAGCGCTGGGC----- 696
||| ||| |||||
Dd 274 HisSer-----SertyrLeuGlYthrAspTyrrProThralamethrProThr 289

Qy 697 -----TCCCTGGAAATCGGGAA-----AACAG 720
||||| :|||
Dd 290 SerProargArgTyrrSerProvalAlaLysaspleuLeuGlyGluGluaspilleProArg 309

Qy 721 GAGAAGAAGTCTTCATCAGCGTGGTAGGCTCCGAGCGCTTGGCTGCAGATTCACG 780
||| :||| ||||| ||||| ||||| |||||
Dd 310 GluProArqIleValIleHisArgGlySerThrglyLeuGlyPheAsnIleValGlv 329


```
QY 781 GCCCCTCCAGAGCCCTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
   |||
   |||
Db 330 GlyGlu---AspGlyGluGlyPheIleLeuAlaGlyGlyProAlaAsp 348
   |||
   |||
QY 832 CTGCTCTGAGGTGGGATTTGAGATAGGGACAGATGTCGAAGTCAATGGCGCTCAG 891
   |||
   |||
Db 349 LeuSerGlyGlu---LeuArgGlyGlyAspGlnIleLeuSerValAsnGlyValAsp 366
   |||
   |||
QY 892 TTTCTCACTGATCACAAGGAGGCTGTAAATGTGCTGAAATAAGCGCGACCTGACC 951
   |||
   |||
Db 367 LeuArgAsnAlaSerHisGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 385
   |||
   |||
QY 952 ATCTCATTTAGTCTGAGCTGGCGGAGCTGTTTCATCAGACACGGGAGCGCTGGCA 1011
   |||
   |||
Db 386 ValThrIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
   |||
   |||
QY 1012 GAGGCGGCGCAGCTGAGCTGAGCGGCGGAGGAGCTTCTCATCCAGAAGCGGCTGCGATG 1071
   |||
   |||
Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
   |||
   |||
QY 1072 GAGTCCAAACAAGATCTCCAGGAGCAGCAGGAG-----ATGGAG 1110
   |||
   |||
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
   |||
   |||
QY 1111 CGSAAAGGAGAAAGAAATGCCAGAGGCGAGAGGAGAAATCAGAGATACCGGAAG 1170
   |||
   |||
Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 458
   |||
   |||
QY 1171 GAGATGGACAGATTTAGAGGAGGAGGAGAGATTT----- 1206
   |||
   |||
Db 459 ValLeuHisValIleAspAlaGlyAspGluGluTrpTrpGlnAlaArgValHisSer 478
   |||
   |||
QY 1207 -----AAGAAGCAA 1215
   |||
   |||
Db 479 AspSerGluThrAspAspIleGlyPheIleProSerLysArgArgValGluArgGlu 498
   |||
   |||
QY 1216 TGGGAAGAAGACTGGGGCTCA-----AAGGAACAGCTACTCTTG 1254
   |||
   |||
Db 499 TrpSerArgLeuTrpGlySerSerGlySerGlnGlyArgGluAspSerValLeuSer 518
   |||
   |||
QY 1255 CCTAAACCATCACT---GCTGAGGTACACCCAGTACCCCTTCGCGAAGCAAGTGATTT 1311
   |||
   |||
Db 519 TyrGluThrValThrGlnMetGluValHisTyrAlaArg----- 531
   |||
   |||
QY 1312 CCGAATATGAGGAGGCTTTGACCCCTACTCTATGTTCACCCAGCAGCATCGG 1371
   |||
   |||
Db 532 -----ProIleIleLeuGlyProThrLysAspArgAla 543
   |||
   |||
QY 1372 GAAGGATGTCCGGCTCTAGCATCAAGAAGGAGGATCTTAGACCTGGCCCTGGGAAG 1431
   |||
   |||
Db 544 -----AsnAspAspLeuLeuSerProSerProAspLys 554
   |||
   |||
QY 1432 CGGTGTGGACTCCCATTTGGGAAGTGGTCTGCTGTATGACGGGGGAGCTGC 1491
   |||
   |||
Db 555 -----PheGlySerCys 558
   |||
   |||
QY 1492 TGAGCGCATGTTGGCATTGTGAAGG---GGACGAGATCATGGCAATCAACGGCAAGAT 1548
   |||
   |||
Db 559 ValPro-----HisThrThrArgProLysArgGluTyr---GluIleAspGlyArg-- 574
   |||
   |||
QY 1549 TGTGACAGACTACCCCTGGCTGAGGCTGACGCTGCCGTGACAGAGGCGCTGGGAATCAGG 1608
   |||
   |||
Db 575 -----AspTyrHisPheValSerArgGluLysMetGluLys----- 587
   |||
   |||
QY 1609 CGGGGACTGATCGACCTTGTGTTGCCCTGTCGCCCCCAAGAGGATGACCATGAGCT 1668
   |||
   |||
Db 588 -----AspIleGln 591
   |||
   |||
QY 1669 GACCTTCTTCTGAGTCCAAAGGGGAACCAAAATTCACCGCTTAGGAACAGT----- 1723
   |||
   |||
Db 591 aHisLysPheIleGluAlaGlyGlnTyrAsnSer---HisLeuTyrGlyThrSerValGln 610
   |||
   |||
```

```
QY 1724 -----GAGTCCGGCCCGCCACCTCGTGAACACAAAGCCTCGGACCAG 1764
   |||
   |||
Db 610 nserValargGluValAlaGluGlnGlyLysHisCysIleLeuAspValSerAlaAsnAl 630
   |||
   |||
QY 1765 CTTGAGAGAGCCACATGACACACACAGATGCATCTTGGGACCTGAATCTATCACC 1824
   |||
   |||
Db 630 aValArgLeuGlnAlaAlaHisLeu-----HisProIleAlaIlePheIleArgPr 648
   |||
   |||
QY 1825 CAGGAATCTCAAACTCCCTTTGGCCCTGAAC-----CAGGCCACAGAT 1866
   |||
   |||
Db 648 oArgSerLeuGlnAsnValLeuGluIleAsnLysArgIleThrGluGlnAlaArgL 668
   |||
   |||
QY 1867 AAGGAACAGCTCGGCGCCACACTTTTGAAG-----GCCAATG 1902
   |||
   |||
Db 668 ysAlaPheAspArgAlaThrLysLeuGluGlnGluPheThrGluCysPheSerAlaIleV 688
   |||
   |||
QY 1903 TGGGAGAAAGGAGCAGCAGCCGCTTTGGGAGAAGATCTCAAGG---ATCCAGACTCTCA 1959
   |||
   |||
Db 688 alGluGlyAspSerPheGluGluIleTyrHisLysValLysArgValIleGluAspLeuS 708
   |||
   |||
QY 1960 TT---CCTTCTCTCTGG 1973
   |||
   |||
Db 708 erGlyProTyrIleTrp 713
   |||
   |||
RESULT 4
T23160
Hypothetical protein K01A6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C:Accession: T23160
R: Cottage, A.
submitted to the EMBL Data Library, January 1996
A: Reference number: Z19701
A: Accession: T23160
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-1012 <WIL>
A: Cross-references: EMBL:Z68750; PIDN:CAA92965.1; GSPDB:GN00022; CESP:K01A6.1
A: Experimental source: clone K01A6
C: Genetics:
A: Gene: CESP:K01A6.1
A: Map position: 4
A: Introns: 107/1; 151/2; 183/2; 270/2; 316/3; 410/3; 628/3; 922/3; 971/3
A: Superfamily: WW repeat homology
F: 131-168/Domain: WW repeat homology <WWR>
Alignment Scores:
Pred. No.: 3,73e-08 Length: 1012
Score: 243.00 Matches: 123
Percent Similarity: 36.47% Conservative: 71
Best Local Similarity: 23.12% Mismatches: 191
Query Match: 6.09% Indels: 147
DB: 2 Gaps: 19
US-09-502-945-5 (1-2162) x T23160 (1-1012)
QY 29 CCAGCTCTCGCAGCGGGGACCCCGAAGAGCGGTCTG-----GCACGACGAG 76
   |||
   |||
Db 574 ProLysThrArgSerArgThrProSerAlaAlaPheArgTyrGlyGluProGlnThrAsn 593
   |||
   |||
QY 77 CAGCTGGACCTGGCCGACGCA-----TGGACC 103
   |||
   |||
Db 594 MetMetAspSerAlaAlaProLeuProValArgSerLysThrProAlaGluArgGlnThr 613
   |||
   |||
QY 104 GAAAGTGGCCCGGAGAATTCCGGCATAAGGTGATTTCTGATTGAAATGATGCAGAGA 163
   |||
   |||
Db 614 SerArgThrGluGluAspGlnAsnValArg-----AsnThrLeuGlnArg 628
   |||
   |||
QY 164 AGGACTATCTCTATGATGTGTCGAATGTACCAACGACCATGGACGCTGGCGTCTCG 223
   |||
   |||
Db 629 Gln-----ProAlaValThrSerGluTrp---GluGlyMetSerSe 641
   |||
   |||
QY 224 TGGGAGACCTGAAGCTGGTCAATCAATGAACCCAGCCGCTGCTGCTCTGTTTGTGCAATTC 283
   |||
   |||
```



```
QY 598 GGCCTGATCCCTCTCTGATGAGCCCTCACTTGGCAGTATGTG 651
   |||
   |||
Db 413 GlyTyrAlaProProAspIleThrAsnSerSerGlnPro-----Val 427
QY 652 GATCAGTTGTGTCGGAATCTGGGGCGGTGCGAGCGCCTGGCTCCCTCT 702
   |||
   |||
Db 428 AspAsnHisValSerProSerSerPheLeuGlyGlnThrProAlaSerProAlaArgTyr 447
QY 703 -----GGAAATCGGGAA-----ACRAGGAGGAAGAGTCTTC 735
   |||
   |||
Db 448 SerProValSerLysAlaValLeuGlyAspAspGluLeuThrArgGluProArgLysVal 467
QY 736 ATCAGCTGGTAGGCTCCCGAGGCTTGGCTGCAGCATTTCCAGCGCCCTCCAGAAG 795
   |||
   |||
Db 468 ValLeuHisArgGlySerThrGlyLeuGlyPheAsnIleValGlyGlyGlu---AspGly 486
QY 796 CTTGCATCTTTATCAGCATGTG-----AAACCTGGCTCCTGTCTGCTGAGGTG 846
   |||
   |||
Db 487 GluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu--- 505
QY 847 GGATTGGAGATAGGGACCATGTCGAAGTCATGCGCTCGACTTCTTAACCTGGAT 906
   |||
   |||
Db 506 ---LeuArgLysGlyAspArgIleIleSerValAsnSerValAspLeuArgAlaAlaSer 524
QY 907 CACAGGAGGCTGTAAATCTGTGTAATAATAGCCGAGCTGACCATCTCCATTTAGCT 966
   |||
   |||
Db 525 HisGluGlnAlaAlaAlaLeuLysAsnAla---GlyGlnAlaValThrIleValAla 543
QY 967 GCAGCTGCCGGGAGCTGTTTCATGACAGCCGGGAGCGGTGGCAGAGCGCGGAG--- 1023
   |||
   |||
Db 544 GlnTyrArgProGluGlySerArgPheGluAlaLysIleHisAspLeuArgGluGln 563
QY 1024 -----CGTGAGCTCGAGCGGAGGAG 1044
   |||
   |||
Db 564 MetMetAsnSerSerIleSerSerGlySerGlySerLeuArgThrSerGlnLys-ArgSe 583
QY 1045 CTTCTCATCGAAGCGGTGGCGATGGAGTCCACAAAGATCCTCC----- 1090
   |||
   |||
Db 583 rLeuTyrValArgAlaLeuPheAspTyrAspLysThrLysAspSerGlyLeuProSerG1 603
QY 1091 -----AGACGACGACGAGATG 1107
   |||
   |||
Db 603 nGlyLeuAsnPheLysPheGlyAspIleLeuHisValIleAsnAlaSerAspAspGluTr 623
QY 1108 G-----AGCGCAAGAGAGAA----- 1123
   |||
   |||
Db 623 pTrpGlnAlaArgGlnValThrProAspGlyGluSerAspGluValGlyValIleProSe 643
QY 1124 -AAGAAATGCCCCAGAGCGACGAGG-----AA 1152
   |||
   |||
Db 643 rLysArgArgValGluLysLysGluArgAlaArgLeuLysThrValLysPheAsnSerLys 663
QY 1153 AATGAGAGATACCGNAGAGATGGACAGATGTAGAGGAGGAAGAGATTTAAGAG 1212
   |||
   |||
Db 663 sThrArgAspLysGlyGlnSerPheAsnAsp-----LysArgLysLysAsnLeuPheSe 681
QY 1213 C-----AATGGGAAGAAGAGACTGGGGCTCAAAGCAAGCACTACTCTTG 1254
   |||
   |||
Db 681 rArgLysPheProPheTyrLysAsnLysAspGlnSerGluGlnGluThrSerAspAla--- 700
QY 1255 CCTAAACCATCACTGCTGAGGTACACCCAGTACCTCCGCAAGCCAAAGTATTCCG 1314
   |||
   |||
Db 701 -----AspGlnHisValThrSerAsnAlaSerAspSerG1 712
QY 1315 GAATATGAGGAGGCTTTGACCCCTACTCTATGTTCCACCCAGCAGCATATGGGAA 1374
   |||
   |||
Db 712 uSerSerTyrArgGlyGlnGluGluTyr---ValLeuSerTyrGluProValAsnGlnG1 731
QY 1375 GGATGTCGGCTCCTACGATCAAGAAGGAGGATCTTTAGACCTCGCCCTCGAAGCGG 1434
   |||
   |||
Db 731 nGluValAsnTyrThrArg-----ProValIleIleLeuGlyPr 744
```

```
QY 1435 TGTGGACTCCCCCATTTGGGAAGGTGCTGTTTCTGCTGTGTATGACCGG---GGACCTGC 1491
   |||
   |||
Db 744 oMetLysAspArgIleAsnAspAspLeuIleSerGluPheProAspLysPheGlySerCy 764
QY 1492 TGAGCGGCATGTTGGCATTTCTGAAGGGGAC---GAGATCATGGCAATCAACGCAAGAT 1548
   |||
   |||
Db 764 sValProHisThrThrArgProLysArgAspTyrGluValAspGlyArgAspTyrHisPh 784
QY 1549 TGTGACAGACTACACCTGCTGAGCTGACCTGCGCTGCAGAAAGCCCTGGAAATCAGGG 1608
   |||
   |||
Db 784 eValThrSerArgGluGlnMetGluLysAspIleGlnGluHisLys---PheIleGluAl 803
QY 1609 CGGGAGCTGGATCGACCTTGTG-----GTTGCCGTCTGCCCCCAAGAGGATGACGA 1662
   |||
   |||
Db 803 aGlyGlnTyrAsnAsnHisLeuTyrGlyThrSerValGlnSerValArgGluValAlaG1 823
QY 1663 TGAGCTGACCTTCTGCTGAAGTCCAAAAGGGAAACCAAAATTCACGCGTTAGGAAACAG 1722
   |||
   |||
Db 823 yLysGlyLysHisCysIleLeuAspValSerGlyAsnAlaIleLysArgLeuGlnIleAl 843
QY 1723 TGAGCTCGGCGCCACCTCGTGAACACAAAGCCCTCGGACC 1762
   |||
   |||
Db 843 aGlnLeuTyrProIleSerIlePheIleLysProLysSer 856
RESULT 6
A46431
tight junction-associated protein ZO-1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A46431
R:Itch, M.; Nagafuchi, A.; Yonemura, S.; Kitani-Yasuda, T.; Tsukita, S.;
J. Cell Biol. 121, 491-502, 1993
A:Title: The 220-kD protein colocalizing with cadherins in non-epithelial cells is id
py.
A:Reference number: A46431; MUID:93252986; PMID:8486731
A:Accession: A46431
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1745 <IT0>
A:Cross-references: GB:D14340; NID:g303709; PIDN:BAA03274.1; PID:d1003784; PID:g30371
A:Experimental source: F9 cells
A:Note: sequence extracted from NCBI backbone (NCBI:131200, NCBIP:131201)
C:Superfamily: guanylate kinase homology; GLGF domain homology
F:27-106/Domain: GLGF domain homology <GLG1>
F:428-498/Domain: GLGF domain homology <GLG3>
F:645-794/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 1.33e-07 Length: 1745
Score: 236.00 Matches: 146
Percent Similarity: 37.63% Conservative: 102
Best Local Similarity: 22.15% Mismatches: 218
Query Match: 5.92% Indels: 193
DB: 2 Gaps: 32
US-09-502-945-5 (1-2162) x A46431 (1-1745)
QY 358 GTGGCTGTGGACCGTCTGCACCGCGAGCCCTCGCTGAGTGTGCTGTGGCTGGAG 417
   |||
   |||
Db 24 ValThrLeuHisArgAlaProGlyPheGlyPheGlyIleAlaIleSerGlyArgAsp 43
QY 418 -----TTTGGCTGTGGG-----CTCTTCATCTCCCACTCCTCAAGGCGGT 459
   |||
   |||
Db 44 AsnProHisPheGlnSerSerGlyGluThrSerIleValIleSerAspValLeuLysGlyGly 63
QY 460 CAGGACAGACCGCTCGGCTCCAGGTAGGGACGAGATCGTCCGGATCAATGATATATCC 519
   |||
   |||
Db 64 ProAlaGluGly---GlnLeuGlnGluAsnAspArgValAlaMetValAsnGlyValSer 82
QY 520 ATCTCTCTCTGATCCCATGAGGAGGTGTCATCAAC----- 552
   |||
   |||
Db 83 MetAspAsnValGluHisAlaPheAlaValGlnGlnLeuArgLysSerGlyLysAsnAla 102
```



```

QY 553 -----CTCATTGCAACCAAGAACTGTGTCATCAAAAGTGAGACATCGGCTGATC 606
Db 103 LysIleThrIleArgArgLysLysValGlnIleProValSerHis----- 118
QY 607 CCGGTGAAAGCTCTCCATGAT---GAGCCCTCACTTGGCAGTATGTGATCAGTTTGTG 663
Db 119 -----ProAspProGluProValSerAspAsnGluAspSerTyrAsp 133
QY 664 TCGGAATCTGGGGCTGCGAGGAGCCCTGGCTCCCTCGGAATTCGGGAAACAAAGAG 723
Db 134 GluGluValHisAspProArgAlaGlyArgGlyAlaLeuAlaAsnArgSerGluLys 153
QY 723 ----- 723
Db 154 SerTrpAlaArgAspArgSerAlaSerArgGluArgSerLeuSerProArgSerAspArg 173
QY 724 -----AAGAAGTCTTTCATCAGCCTGTAGGCTCCCGA 756
Db 174 ArgSerValAlaSerSerGlnProAlaLysProThrLysValThrLeuValLysSerArg 193
QY 757 -----GGCTTGGCTGCAGCATTTCCAGCGGCCCATCCAGACCTGGC 801
Db 194 LysAsnGluGluTyrGlyLeu-----ArgProAla 203
QY 802 -----ATCTTATCAGCCATGTGAACCTGGCTCCCTGTCTGCTGAGGTGGGA---TTG 852
Db 204 SerHisIlePheValLysGluIleSerClnAspSerLeuAlaAlaArgAspGlyAspIle 223
QY 853 GAGATAGGGACCAAGTTGTCGAGTCAATGGCTGCGACTTCTTAACCTGGATCACAAG 912
Db 224 GlnGluGlyAspValValLeuLysIleAsnGlyThrValThrGluAsnMetSerLeuThr 243
QY 913 GAGCTGTAATGTGCTCAAAATAGCCGACCTGACCATCTCCATGTAGCTGCAGCT 972
Db 244 AspAlaLysThrLeuIleGluArgSerLysGlyLysLeuLysMetValValGlnArgAsp 263
QY 973 GCGCGGAGCTGTTCATG----- 990
Db 264 GluArgAlaThrLeuLeuAsnValProAspLeuSerAspSerIleHisSerAlaAsnAla 283
QY 991 ACAGCCGGGAGCGGTGGCAGAGCGCGGCGAG-----CGTGAG 1029
Db 284 SerGluArgAspAspIleSerGluIleGlnSerLeuAlaSerAspHisSerGlyArgSer 303
QY 1030 CTGCAGCGGAGGAGCTCTCATCGACAGCGGCTG----- 1065
Db 304 HisAspArgProProArgArgSerGlnSerArgSerProAspGlnArgSerGluProSer 323
QY 1066 -----GCGATGGAGTCCCAACCAAGATCCTCCAGGAGCAGCAGAG 1104
Db 324 AspHisSerThrGlnSerProGlnGlnProSerAsnGlySerLeuArgSerArgGluGlu 343
QY 1105 ATGAGCGGCAAGAGGAAAGAAATTTCCCAAGAGGACAGCAGGAAATAGAGATAC 1164
Db 344 GluArgMetSerLysProGlyAlaIleSerThrProValLysHisValAspAspHisPro 363
QY 1165 CGGAGGAGATGGACAGATT---GTACAGGAGGAAGAGAGTTTACAGACCAATGGCAA 1221
Db 364 ProLysAlaValGluGluValThrValGluLysAsnGluLysGlnThrProThrLeuPro 383
QY 1222 GAAGACTGGGGCTCAAAGGAACAGCTACTCTTGCTATAAACCATCACTGCTGAGGTACAC 1281
Db 384 Glu-----ProLysProValTyrAlaGlnValGly 393
QY 1282 -----CCAGTACCCCTTCGCAAGCCAAAGTGTTCGGAATAATGA 1323
Db 394 GlnProAspValAspLeuProVal-----SerPro-SerAsp----- 405
QY 1324 GGAAGGCTTTGACCCCTACTCTATGTTCACCCACAGACATCATGGGAGGATGTCG 1383
Db 406 ---GlyAlaLeuProAsnSerAlaHis---GluAspGlyIleLeuArgProSerMetLys 423
QY 1384 GCTCTACGATCAAGAGGAGGATCTTAGACCTGGCCCTGGAAGCGGTGTGGACTC 1443

```

```

Db 423 sLeuValLysPheArgLysGlyAspSerValGlyLeuArgLeuAlaGlyGlyAsnAsp-- 442
QY 1444 CCCATTGGGAAGGTGCTGCTTTCTGCTGTATGACGGGAGCTGCTGAGCGGCATGG 1503
Db 443 -----ValGly---IlePheValAlaGlyValLeuGluAspSerProAlaAlaLysGlu-- 459
QY 1504 TGGCATTTGTAAGGGACGAGATCATGGAATCAACGGCAAGATTGTGACAGACTACAC 1563
Db 460 -GlyLeuGluGluGlyAspGlnIleLeuArgValAsnAsnValAspPheThrAsnIleI 479
QY 1564 CCTGGCTGAGGCTGACGCTGCCCTGTCAGAGAGGCTGGAATCAGGCGGGGACTGGATCGA 1623
Db 479 eArgGluGluAlaValLeuPheLeu-----LeuAs 489
QY 1624 CTTGTGTGTTGCCGCTGCTGCCCCCAAGAGTATGACGATGAGCTGACCTTCTTGTCTGAA 1683
Db 489 pLeu-----ProLys-----GlyGluGluValThrIleLeuAlaG 501
QY 1684 GTCCAAAGGGGAACCAAAATTCACGCGTTAGGAACAGTGCCTCGGCCCCACCTCGT 1743
Db 501 nLysLysLys-----AspValTyrArgArgIleVa 511
QY 1744 GAACACAAAGCCTCGGACACGCTTGAG---AGAGGCCACATGACACACAGATGGCA 1800
Db 511 lGluSerAspValGlyAspSerPheTyrIleArgThrHisPheGluTyrGluLysGluSe 531
QY 1801 TCCTTGGGACCTGAATCTATCACCAGGAATCTCAAACTCCCTTTGGCCCTCAACAGGG 1860
Db 531 rProTyrGly-----LeuSerPheAsnLysG 540
QY 1861 CCGATTAAGGAACAGCTCGGCGCCACTTTTGAAGGCCAATGTGGAGGAAGGAGGAGCAGC 1920
Db 540 yGluValPheArgValValAspThrLeuTyrAsnGlyLysLeu---GlySerTrpLeuAl 559
QY 1921 CAGCGCTTTCGGAGAGATCTCAAGGATCCAGACTCTCATCTCTCTTCTCTCTCTCTCT 1969
Db 559 alleArgIleGlyLysAsnHisLysGluValGluArgGlyIleIleProAsnLysAsnAr 579
QY 1970 -----CTGCGCCAGTGAATTTGGTCTCTCCCA-----GCTTTGGGGGAC 2008
Db 579 alaGluGlnLeuAlaSerValGlnTyrThrLeuProLysThrAlaGlyGlyAsp 597

```

RESULT 7

T09599

postsynaptic density protein 95 - human

C:Species: Homo sapiens (man)

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000

C:Accession: T09599

R:Stathakis, D.G.; Hoover, K.H.; You, Z.; Bryant, P.J.

submitted to the EMBL Data Library, July 1998

A:Reference number: Z16761

A:Accession: T09599

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-767 <STA>

A:Cross-references: EMBL:U83192; NID:g3318652; PID:g3318653

A:Experimental source: mammary

C:Genetics:

A:Gene: PSD95

C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h

F:208-286/Domain: GLGF domain homology <GLG>

F:478-536/Domain: SH3 homology <SH3>

F:578-755/Domain: guanylate kinase homology <GKI>

Alignment Scores:

Pred. No.: 1,24e-07

Score: 234.50

Percent Similarity: 38.10%

Best Local Similarity: 22.12%

Query Match: 5.88%

Indels: 2

Gaps: 34

Length: 767

Matches: 144

Conservative: 104

Mismatch: 210

Indels: 193

Gaps: 34

A:Molecule type: mRNA

A:Residues: 1-724 <CHO>

A:Cross-references: GB:M95853; NID:G206454; PIDN:AAA1971.1; PID:G206455

A:Experimental source: brain

C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase hom

F:70-148/Domain: GLGF domain homology <GLG1>

F:165-243/Domain: GLGF domain homology <GLG2>

F:435-493/Domain: SH3 homology <SH3>

F:535-712/Domain: guanylate kinase homology <GKI>

Alignment Scores:

Pred. No.:	1-66e-07	Length:	724
Score:	232.50	Matches:	144
Percent Similarity:	38.10%	Conservative:	104
Best Local Similarity:	22.12%	Mismatches:	210
Query Match:	5.83%	Indels:	193
DB:	2	Gaps:	34

US-09-502-945-5 (1-2162) x JH0800 (1-724)

```

QY 340 TCAGGAAGCTGAAGAGGTGGCTGTGGACCTGTGCACCCGAAAGGCTCGGCTGAGT 399
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 155 AlaGluValMetGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 173
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 400 GTGCGGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 444
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 174 IleAlaGlyValGlyValGlyValGlyValGlyValGlyValGlyValGlyVal 193
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 445 CTCATCAAGAGGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 501
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 194 IleIleGluGlyValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaLeu 213
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 502 CGATCAATGATATTCATCTCCCTCTGCTACCCAGGAGGAGGAGGAGGAGGAGGAG 561
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 214 AlaValAsnSerValGlyLeuGluLeuAspValMetHisGluAspAlaValAlaLeu 233
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 562 ---ACCAAGAAACTGTGTCATCAAGAGT---AGACATCGGCTGATCCCC 609
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 234 AsnThrTyrAspValValTyrLeuLeuValValAlaValProSerAsnAlaTyrLeu 253
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 610 GTGAAAGCTCTCTGATGAGCCCTCACTTGG---CAGTATGTGATCAGTTTGTGCG 666
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 254 SerTyrAlaProAspIleThrSerTyrSerGlnHisLeuAspAsnGluIleSer 273
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 667 GAATCGGGGCGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 696
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 274 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 289
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 697 -----TCCCTCGGAATCGGAA-----AACCAAG 720
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 290 SerProArgTyrSerProValAlaValAspLeuLeuGlyGluValAspIleProArg 309
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 721 GAGAGAAAGTCTTCATCAGCTGTGAGGCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 310 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 329
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 781 GGGCCCATCCAGAGCCCTGATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 330 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 348
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 832 CTGTCTGTGAGTGGGATTCAGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 891
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 349 LeuSerGlyGlu-----LeuArgGlyGlyAspGlnIleLeuSerValAsnGlyValAsp 366
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 892 TTCTCTAACCTGATACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 951
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 367 LeuArgAsnAlaSerHisGluGluAlaAlaIleAlaLeuLeuLeuLeuLeuLeuLeu 385
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 952 ATCTCCATGTGAGTGCAGCTGGCGGAGCTGTTCATGACAGACCCGGGAGCGGCTGCCA 1011
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 386 ValThrIleIleAlaGln-----TyrLeuProGluGluTyrSerArgPhe 400
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1012 GAGCGGGCGGCGGTGAGTGCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1071
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||

```

```

Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1072 GAGTCCCAACAAGATCTCCAGGAGCAGCAGGAG-----ATGGAG 1110
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1111 CGGCAAGAGGAGAAAGAAATTCGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1170
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 458
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1171 GAGATGGACACAGATTTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1206
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 459 ValLeuHisValIleAspAlaGlyAspGluGluTyrTrpGlnAlaArgValHisSer 478
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1207 -----AAGAAGCAA 1215
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 479 AspSerGluThrAspAspIleGlyPheIleProSerLysArgValGluArgGlu 498
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1216 TGGGAA-----GAAGACTGGGCTCA-----AAGGAA 1242
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 499 TrpSerArgLeuLysAlaLysAspTrpGlySerSerGlySerGlnGlyArgGluAsp 518
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1243 CAGCTACTCTTTCCTAAACCATCACT---GCTGAGGTACACCCAGTACCCCTTCGCAAG 1299
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 519 SerValLeuSerTyrGluThrValThrGlnMetGluValHisTyrAlaArg----- 535
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1300 CCAAGTGATTTCCGGAAATATGAGGAAGGCTTTGACCCCTACTCTATGTTACCCGAGA 1359
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 536 -----ProIleIleLeuGlyProThr 543
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1360 GCAGATCATGGGAAGGATGTCGGCTCCTACGCATCAAGAGGAGGAGGAGGAGGAGGAG 1419
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 544 LysAspArgAla-----AsnAspAspLeuLeuSerGlu 554
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1420 GGCCTTGAAGGCGGTGTGGACTCCCGCATTTGGGAAGGTGCTGCTTCTGCTGTATGA 1479
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 555 PheProAspLys----- 558
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1480 GCGGGAGCTGCTGAGCGGCATGTGGCATGTGTGAAAG---GGACGAGATCATGGCAAT 1536
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 559 PheGlySerCysValPro-----HisThrThrArgProLysArgGluTyr-GluI 575
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1537 CAACGGCAAGATTGTGACAGACTACACCTCGGTGAGGCTGACCTGCCCTGAGAGAGG 1596
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 575 eAspGlyArg-----AspTyrHisPheValSerSerArgGluLysMetGluLys-- 591
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1597 CTGAATCAGGGCGGAGCTGGATCGACCTTGTGTCGCTCTGCTGCCCTCCCAAGGAGTA 1656
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 591 ----- 591
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1657 TGAGCATGAGCTGACCTCTTGTGTAAGTCCAAAGGGGAAACCAATTCACCGTTAGG 1716
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 592 -AspIleGlnAlaHisLysPheIleGluAlaGlyIleTyrAsnSer---HisLeuTyrGI 610
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1717 AAACAGT-----GAGCTCCGGCCCCCACCCTCTGTACACACAAA 1752
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 610 YThrSerValGlnSerValArgGluValAlaGluGlnGlyHisCysIleLeuAspVa 630
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1753 GCCTCGACCGAGCTTGTGAGAGAGCCACATGACACACACAGATGGCATCTCTGGGACCT 1812
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 630 1SerAlaAsnAlaValArgLeuGlnAlaAlaHisLeu-----HisProIleAlaII 648
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1813 GANTCTATCACCGAGGATCTCAATCCCTTTGGCCCTGAAC----- 1855
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 648 ePheIleArgProArgSerLeuGluAsnValLeuGluIleAsnLysArgIleThrGluGI 668
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1856 -CAGGGCCAGATAGGAGACAGCTCGGCGCCACTTTTGTGAAG----- 1895
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 668 uGln-AlaArgLysAlaPheAspArgAlaThrLysLeuGluGlnGluPheThrGluCysP 688
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
QY 1896 -----GCCAATGTGGAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1947
    ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||

```


Db 688 heSerAlaIleValGluGlyAspSerPheGluGluIleTyrHisLysValLysArgValI 708
Qy 1948 TCCAGACTCTCATT---CCTTTCCTCGG 1973
Db 708 leGluAspLeuSerGlyProTyrIleTrp 717
RESULT 9
138757
homolog of Drosophila discs large protein, isoform 1 --human
C:Species: Homo sapiens (man)
C:Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
C:Accession: I38757
R:Lue, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.
Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila disc
A:Reference number: I38756; MUID:95024052; PMID:7937897
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-904 <RES>
A:Cross-references: EMBL:U13897; NID:g558437; PIDN:AAA50599.1; PID:g558438
C:Genetics:
A:Gene: GDB:DLG1
A:Cross-references: GDB:393278; OMIM:601014
A:Map position: 3q29-3q29
C:Superfamily: discs-large tumor suppressor; G1GF domain homology; guanylate kinase hom
C:Keywords: alternative splicing; duplication
F:229-307/Domain: G1GF domain homology <G1GF>
F:324-402/Domain: G1GF domain homology <G1GF>
F:588-646/Domain: SH3 homology <SH3>
F:715-892/Domain: guanylate kinase homology <GKI>
Alignment Scores:
Pred. No.: 1.94e-07 Length: 904
Score: 232.00 Matches: 82
Percent Similarity: 47.02% Conservative: 52
Best Local Similarity: 28.77% Mismatches: 103
Query Match: 5.82% Indels: 48
DB: 2 Gaps: 15
US-09-502-945-5 (1-2162) x I38757 (1-904)
Qy 340 TCCAGGAAGCTGAAGAGGTGGCTGGACCGCTGTCACCCGAGGCGCTGGCGCTGAGT 399
Db 314 SerGluLysIleMetGluIleLysLeuLys---GlyProLysGlyLeuPheSer 332
Qy 400 GTGCGTGTGGC-----CTGGAGTTGGCTGTGGCGCTTCTCATCTCCAC 444
Db 333 IleAlaGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 352
Qy 445 CTCATCAAGGGCGTCAGCAGCAGCGTCGGG---CTCCAGTAGGGGAGCAGATCGTC 501
Db 353 IleIleGluGlyGlyAlaAlaHisLysAspGlyLysLeuGlnIleGlyAspLysLeuLeu 372
Qy 502 CGGATCAATGGATATTCATCCCTCTCCCTGACCATGAGGAGGTGATCAACCTCATTCGA 561
Db 373 AlaValAsnValCysLeuGluGluValThrHisGluAlaValThrAlaLeuLys 392
Qy 562 ---ACCAAGAAACTGTGTCCATCAAGTG-----AGACACATC 597
Db 393 AsnThrSerAspPheValTyrLeuLysValAlaLysProThrSerMetTyrMetAsnAsp 412
Qy 598 GGCCTGATCCC-----GTGAAAGCTCTCTGATGAGCCCTCCTACTTGGCAGTATG 651
Db 413 GlyTyrAlaProAspIleThrAsnSerSerSerGlnPro-----Val 427
Qy 652 GATCAGTTGTGCGGAATCTGGGCGCGAGGAGCGCTGGCTCCCTC----- 702
Db 428 AspAsnHisValSerProSerPheLeuGlyGlnThrProAlaSerProAlaArgTyr 447
Qy 703 -----GGAAATCGGGA---AACAGGAGGAAGAGTCTTTC 735
Db 448 SerProValSerLysAlaValLeuGlyAspGluIleThrArgGluProArgLysVal 467

Qy 736 ATCAGCCTGGTAGGCTCCGAGCGCTTGGCTGCAGCATTTCAGCGGGCCCCCACCAGAAG 795
Db 468 ValLeuHisArgGlySerThrGlyLeuGlyPheAsnIleValGlyGlu---AspGly 486
Qy 796 CTGCGCATCTTATCAGCCATGTG-----AAACCTGGCTCCCTGTCTGCTGAGGTG 846
Db 487 GluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGlu--- 505
Qy 847 GGATTGGAGTAGGCGACCATGTCGAAGTCAGTGGCGTGCAGCTTCTTAACCTGGAT 906
Db 506 ---LeuArgLysGlyAspArgIleIleSerValAsnSerValAspLeuArgAlaAlaSer 524
Qy 907 CACAGGAGGCTGTAATGTGCTGAAAATAGCCGAGCGCTGACCATCTCCATGTAGCT 966
Db 525 HisGluAlaAlaAlaAlaLeuLysAsnAla---GlyGlnAlaValThrIleValAla 543
Qy 967 GCAGCTGGCGGGAGCTGTTTCATGACAGACGGGAGCGGCTGCAGAGCGGCGGAGCGT 1026
Db 544 Gln-----TyrArgProGluGluTyrSerArgPheGluAlaLysIleHis 558
Qy 1027 GAGCTGAGCGGCGAGGAGCTTCTCAGCAGAGCGGCTGCGGATGGAGTCCAAACAGATC 1086
Db 559 AspLeuArgGluGln-----MetMetAsnSerSerIleSerGlySerGlySerLeu 576
Qy 1087 CTCAGGAGCAGCAG 1101
Db 577 ArgThrSerGlnLys 581
RESULT 10
S71625
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 13 - mouse
N:Alternate names: epidermal growth factor-binding protein; serine proteinase
C:Species: Mus musculus (house mouse)
C:Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 21-Jun-2002
C:Accession: S71625; S67987; I81210; I81209; S40290
R:Chida, D.; Kume, T.; Mukoyama, Y.; Tabata, S.; Nomura, N.; Thomas, M.L.; Watanabe,
FEBS Lett. 358, 233-239, 1995
A:Title: Characterization of a protein tyrosine phosphatase (RIP) expressed at a very
A:Reference number: S71625; MUID:95145716; PMID:7843407
A:Accession: S71625
A:Molecule type: mRNA
A:Residues: 1-2450 <CHI>
A:Cross-references: EMBL:D83966; NID:g1232103; PIDN:BAAL2158.1; PID:g1232104
A:Experimental source: strain DBA/2; Cell line MEL 745A
R:Wolf, B.B.; Brown, M.D.
FEBS Lett. 376, 177-180, 1995
A:Title: Epidermal growth factor-binding protein activates soluble and receptor-bound
A:Reference number: S67987; MUID:96105375; PMID:7498536
A:Accession: S67987
A:Molecule type: protein
A:Residues: 1098-1102 <WOL>
A:Experimental source: submaxillary glands
R:Sato, T.; Irie, S.; Kitada, S.; Reed, J.C.
Science 268, 411-415, 1995
A:Title: FAP-1: a protein tyrosine phosphatase that associates with Fas.
A:Reference number: I59595; MUID:95232528; PMID:7536343
A:Accession: I81210
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1329-1354, 'K', 1356-1447, 'R', 1449-1454 <RES>
A:Cross-references: GB:I34582; NID:g806297; PIDN:AAC42056.1; PID:g806298
A:Accession: I81209
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1338-1354, 'K', 1356-1447, 'R', 1449-1454 <RE2>
A:Cross-references: GB:I34581; NID:g806295; PIDN:AAC42055.1; PID:g806296
R:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.
submitted to the EMBL Data Library, June 1993
A:Description: Assessment of the expression levels of murine protein-tyrosine phosph
A:Reference number: S40280
A:Accession: S40290
A:Molecule type: mRNA

A:Residues: 2266-2372 <HEN>
 A:Cross-references: EMBL:Z23059; NID:g438155; PIDN:CAAR0594.1; PID:g438156
 C:Genetics:
 A:Gene: Ptpn13
 A:Map position: 5
 C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology;
 F:566-860/Domain: phosphoric monoester hydrolase; transmembrane protein; tyros
 F:1089-1165/Domain: protein 4.1 membrane-binding domain homology <B41>
 F:1361-1437/Domain: GLGF domain homology <GLG1>
 F:1495-1574/Domain: GLGF domain homology <GLG2>
 F:1769-1840/Domain: GLGF domain homology <GLG3>
 F:1863-1937/Domain: GLGF domain homology <GLG4>
 F:2203-2422/Domain: GLGF domain homology <GLG5>
 F:2374/Active site: Cys (phosphotyrosine intermediate) #status predicted
 F:2380/Binding site: substrate phosphate (Arg) #status predicted

Alignment Scores:

Pred. No.:	6.02e-07	Length:	2450
Score:	227.00	Matches:	146
Percent Similarity:	32.23%	Conservative:	106
Best Local Similarity:	18.67%	Mismatches:	192
Query Match:	5.69%	Indels:	338
DB:	2	Gaps:	30

US-09-502-945-5 (1-2162) x S71625 (1-2450)

QY	308	AGRTGGAAATATGATCAGCTGACCCCGCGCTCCAGGAAGCTGAAGGAGTGGCTGCG	367
Db	1070	ArgTrpSerIleValSer--SerProGluArgGluIleThrLeuValAsnLeuLysA	1089
QY	368	ACCGTCTCCACCCGAGGCTCGGCTGAGTGTGCGTGGCTGGAGTTGGC----	423
Db	1089	spProLysHis-----GlyLeuGlyPheGlnIleIleGlyGlyLysMetGlyArgL	1107
QY	424	-----TGTGGCTCTTCATCTCCACCTCATCAAGGCGGTGAGGAGGCTGGG--	477
Db	1107	euAspLeuGlyValPheIleSerAlaValThrProGlyGlyProAlaAspLeuAspGlyC	1127
QY	478	-----CTCCAGGTAGGGAGAGATCGTCGGATCATGATATTCATCTCCCTGACCC	535
Db	1127	ysLeuLysProGlyAspArgLeuIleSerValAsnSerValSerLeuGluGlyValSerH	1147
QY	536	ATGAGGAGGTCAACCTCATTCGA-----	561
Db	1147	isHisAlaAlaValAspIleLeuGlnAsnAlaProGluAspValThrLeuValIleSerG	1167
QY	561	-----	561
Db	1167	lnProLysGluLysProSerLysValProSerThrProValHisPheAlaAsnGlyMetL	1187
QY	562	-----ACCAAGAAACTGTGTCATCAAA-----	585
Db	1187	ysSerTyrThrLysLysProAlaTyrMetGlnAspSerAlaMetAspProSerGluAspG	1207
QY	586	-----CTGACACATCGGCTGATCCCGTGAAAGCTCTCTG	625
Db	1207	lnProTrpProArgGlyThrLeuArgHisIle-----ProG	1219
QY	626	ATGAGCCCTCACTGGCAGATGTGGATCAGTTGTCTCGAATCTGGGGCGCTGGCA-	684
Db	1219	luSerProPhe-----GlyLeuSerGlyGlyLeuArg	1230
QY	685	-----GGCAGCTGGCTCCCTGGAATCGG-----GAAACCAAGG	721
Db	1230	luGlySerLeuSerSerGlnAspSerArgThrGluSerAlaSerLeuSerGlnSerGlnV	1250
QY	722	AGAGAGGCTTTCATCAGCTGGTAGCTCCCGAGCG-----	759
Db	1250	alaAsnGlyPhePheAlaSerHisLeuGlyAspArgGlyTrpGlnGluProGlnHisSers	1270
QY	759	-----	759

Db	1270	erProSerProSerValThrThrLysValAsnGluLysThrPheSerAspSerAsnArgS	1290
QY	759	-----	759
Db	1290	erLysAlaLysArgArgGlyIleSerAspLeuIleGluHisLeuAspCysAlaAspSerA	1310
QY	759	-----	759
Db	1310	spLysAspAspSerThrTyrThrSerSerGlnAspHisGlnThrSerLysGlnGluProS	1330
QY	759	-----	759
Db	1330	erSerSerLeuSerThrSerAsnLysThrSerPheProThrSerSerAlaSerProProL	1350
QY	760	-----CTTGGCTGCAGCA	772
Db	1350	ysProGlyAspThrPheGluValGluLeuAlaLysThrAspGlySerLeuGlyIleSerV	1370
QY	773	TTTCCAGCGGC-----CCCATCCAGAGCCCTGGCATCTTTATCAGCCATCTGAAC	823
Db	1370	alThrGlyGlyValAsnThrSerValArgHisGlyGlyIleTyrValIleAlaIleleP	1390
QY	824	CT----GGCTCCCTGCTGCTGAGTGGATGGAGATAGGGACAGATTGTCCGAAGTCA	880
Db	1390	roLysGlyAlaAlaGluSerAspGlyArgIleHisLysGlyAspArgValLeuAlaValA	1410
QY	881	ATGGCGTCGACTTCTTAACCTGGATCACAAGGAGCTGTAATGTCTGCTGAAATAATGCC	940
Db	1410	snGlyValSerLeuGluGlyAlaThrHisLysGlnAlaValGluThrLeuArgAsnThrG	1430
QY	941	GCAGCTTGACCATCTCCATTGTCAGCTGGCGGGAGCTGTTTTCATCAGCAGACCGGG	1000
Db	1430	lyGlnVal---ValHisLeuLeuLeuLysGlyGln-----ValProThrSerArg	1447
QY	1001	AGCGGTGGCAGAGCGCGGCGGCTGAGCTGAGCGGCGAGGAGCTTTCATGCAGAAAGC	1060
Db	1447	luGlnAspProAlaGlyProGlnSerPro-----	1456
QY	1061	GGCTGGCGATGGAGTCCCAACAAGATCTCCAGGAGCAGGAGGATGGAGCGCAAGGA	1120
Db	1457	-----ProProAspGlnAspAlaGlnArgGlnAlaP	1467
QY	1121	GAAAGAAATTTGCCAGAGGCGAGAGAGATAACCGGAGAGATGGAAC	1180
Db	1467	roGluLysValAlaLysGlnThrPro-----HisValLysAspTyrSerP	1482
QY	1181	AGATTGTAGAGGAG-----GAAGAGAAGTTTAAAGAAGCAATGGGAAGAGACTGGG	1231
Db	1482	heValThrGluAspAsnThrPheGluValLysLeuPheLys-AsnSerSerGlyLeuGly	1501
QY	1232	GCTCAAGGAGACAGCTACTCTGCCTAAACCATCATCTGCTGAGGTACACCCAGTACCCC	1291
Db	1502	-----PheSer	1503
QY	1292	TTCCGAAGCCAAAGTGATTTCCGGAATATGAGGAAGCTTTTGACCCCTACTCTATGTT	1351
Db	1504	PheSerArgGluAspAsn-----	1510
QY	1352	ACCCAGAGCAGATCATCGGGAAGGATGTCGGCTCTACGCATCAAGAAGAGGATGCC	1411
Db	1511	IleProGluGlnIleAsnGlySer-----IleValArgValLysLys-----	1524
QY	1412	TTAGACCTGGCCCTGGGAAGCGGTGTGGACTCCCCCATTTGGGAAGGTGGTCTGCTGCT	1471
Db	1524	-----	1524
QY	1472	GTGTATGAGCGGGAGCTGCTGAGCGCATGTGTGGCATTGTGAAGGGGAGGAGATCATG	1531
Db	1525	LeuPheProGlyGlnProAlaAlaGluSerGlyLysIleAspValGlyAspValIleLeu	1544
QY	1532	GCAATCAACGGCAAGATTGTGCAGACACTACCCCTGGCTGAGGCTGACGCTGCCCTGCAG	1591
Db	1545	LysValAsnGlyAlaProLeuLysGlyLeuSerGlnGlnAspValIleSerAlaLeuArg	1564

Db	279	-----ArgArgValValGluHisThrGlyValIle	288
QY	331	CCCCGGCGTCCAGGAAG-----CTGAAGAGAGTGGCTGTGACCGTGTGCACCCCGAA	384
Db	289	AspAspHisGlyArgLysTrpGluLeuGluAsnIleValLeuGluLysGlyHis---Thr	307
QY	385	GGCTTCGGCTGAGTGTGGTGGGCTG-----CAGTTTGGCTCTGGG	429
Db	308	GlyLeuGlyPheSerIleThrGlyGlyMetAspGlnProThrGluAspGlyAspThrSer	327
QY	430	CTCTTCATCTCCACCTCATCAAGAGCGGTGACGACAGCGTCGGG---CTCCAGGTA	486
Db	328	IleTyrValThrAsnIleIleGluGlyAlaAlaLeuAlaAspGlyArgMetArgLys	347
QY	487	GGGGACGAGATCGTCCGATCAATGGATATTCATCTCTCTGTACCCATGAGGAGGTC	546
Db	348	AsnAspIleIleThrAlaValAsnAsnThrAsnCysGluAsnValLysHisGluValAla	367
QY	547	ATCAACCTCATCTGAACCAAGAAACCTGTG-----TCCATCAAGTGGAGA	591
Db	368	ValAsnAlaLeuLysSerSerGlyAsnValSerLeuSerLeuLysArgArgLysAsp	387
QY	591	-----	591
Db	388	GluAlaPheLeuProIleGlyGlyAsnPheGlySerThrSerTyrLeuArgSerGly	407
QY	592	-----CACATCGGCTGATCCCC	609
Db	408	ValThrProSerValSerAlaGlyAsnLeuGlnHisAlaIleHisSerProSerAlaPro	427
QY	610	GTGAAAGCTCTCCTGTATGAGCCCTCACCTGTGCAGTATGTGCATGCTTGTGTCGAA	669
Db	428	IleHisProProProProValHisHisGlySerLeuSerGlnLeu-----	444
QY	670	TCTGGGCGTGCAGACGACCTGGGCTCCCTCGGAATCGGAAACAGAGGAAGAG	729
Db	445	-----SerValGlyGln-----TyrArgSerThrArgProAsnThr	456
QY	730	GTCTTCATCAGCTGGTA---GGCTCCGAGCGCTGGCTGCAGCATTTCCAGCGGCC	786
Db	457	SerValIleAspLeuValLysGlyAlaArgGlyLeuGlyPheSerIleAlaGlyGln	476
QY	787	ATCCAGAA-----CCTGGATCTTTATCAGCCAT---GTGAAACCTGGC	828
Db	477	GlyAsnGluHisValLysGlyAspThrAspIleTyrValThrLysIleIleGluGly	496
QY	829	TCCCTGTCTGCTAGGTGGGATTGGAGATGAGGACACCATGTCCGATCAATGGCGTC	888
Db	497	AlaAlaGluLeuAspGlyArgLeuArgValGlyAspLysIleLeuGluValAspHisHis	516
QY	889	GACTTCTCACTCGGATCACAGGAGGCTGTAATGTCTGAAAATAGCCGAGCGCTG	948
Db	517	SerLeuIleAsnThrThrHisGluAsnAlaValAsnValLeuLysAsnThrGlyAsnArg	536
QY	949	ACCATCTCCATTGTAGCTGCGTGGCGGAGCTGTTTCATCACAGCCGGAGCGGCTG	1008
Db	537	Val-----ArgLeuLeuIle-GlnGlnGly-----	544
QY	1009	GCAGGCGCGGCGAGCTGAGCTGCAGCGGAGGAGCTTCTCATCCAGACGGCTGGC	1068
Db	545	-----ThrGlyAlaIlePheAsnAspSerAlaSerGlnGly	556
QY	1069	ATGGAGTCCCAACAGATCTCCAGGAGCAGCAGGAGATGGAGCGCAAGGAGAAAGAA	1128
Db	556	nPheMetProThr-----	560
QY	1129	ATTGCCAGAGGACGACGAGGAGAAATGAGAGATACCGGAAGGAGATGGAACAGATTGTA	1188
Db	561	-ThrProIleLeuArgProSerSerValGlnAsp-----TyrAsnArgSerGly	576
QY	1189	GAGGAGAAAGAGATTTAAGAGCAATCGGGAAGAAGACTGGGCTCAAGGAACACGTA	1248
Db	576	nMetGlySerGlnSerHisLeuSerTyrGly-----GlyProLeuAsnThrSerTy	593


```
QY 1249 CTCTTGCTAAACCATCACTGCTGAGGTACACCCAGTACACCCCTTCGCAAGCCAAAGTGA 1308
    ||||
Db 593 rSer-----SerGlnAlaProIleAlaIleProLeuGI 604
QY 1309 TTTCCGGAATATGAGGAGGCTTTTGACCCCTACTCTATGTTCCACCCAGCAGATCAT 1368
    ||||
Db 604 uProArg-----ProValGlnLeu-- 610
QY 1369 GGGGAGGATGTCGGGCTCCTAGCATCAAGAAGGAGGATCCTTAGACCTGGCCCTGGA 1428
    ||||
Db 611 -----ValLysGlyGlnAsnGlyLeuGlyPheAsnIleVa 622
QY 1429 AGCGGTGTGAGTCCCTCCCATTTGGGAAGTGGTCTTCTGCTGTATGAGCGGGAGC 1488
    ||||
Db 622 lGlyGlyGluAspAsnGlu-----ProIleTyrIleSerPheValLeuProGlyGlyVa 640
QY 1489 TGTGAGCGGCATGTGGCTGTGAAAGGAGGAGCATGTCGCAATCAACGGCAGAT 1548
    ||||
Db 640 lAlaAspLeuSerGlyAsnValLysThrGlyAspValLeuGluValAsnGlyValVa 660
QY 1549 TGTGACAGACTACACCTGGCTGAGGTGAGCTGCGCTGCAGAGCCCTGGAATCAGG 1608
    ||||
Db 660 lLeuArgAsnAlaThrHisLysGluAlaAlaGluAlaLeuArgAsnAla----- 676
QY 1609 CGGGAGTGGATCCACCTTGTGTTGCTGCGTCTGCCCTCCCAAGGAGTATGACGATGAGCT 1668
    ||||
Db 677 -GlyAsnProValTyrLeuThrLeuGlnTyr---ArgProGlnGluTyrGln----- 692
QY 1669 GACCTCTTGTGTAAGTCCAAAGGGGAAACCAATACACCGTGTAGGAACAGTGAGCT 1728
    ||||
Db 693 -----IlePheGluSerLys-----IleGluLysLeuArgAsnAspValI 706
QY 1729 CCGGCCCCACCTGTGACACAAACGCTCGACCCAGC 1765
    ||||
Db 706 eAlaGlnSerArgMetGlyThrLeuSerArgLysSer 718

RESULT 13
A47747
tight junction protein ZO-1 - human
C:Species: Homo sapiens (man)
C:Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
C:Accession: A47747
R:Willott, E.; Balda, M.S.; Fanning, A.S.; Jameson, B.; Van Itallie, C.; Anderson, J.M.
Proc. Natl. Acad. Sci. U.S.A. 90, 7834-7838, 1993
A:Title: The tight junction protein ZO-1 is homologous to the Drosophila discs-large tum
A:Reference number: A47747; MUID:93361541; PMID:8395056
A:Accession: A47747
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1736 <WIL>
A:Cross-references: GB:L14837; NID:g292937; PIDN:AAA02891.1; PID:g292938
C:Superfamily: guanylate kinase homology; GLGF domain homology
C:Keywords: alternative splicing; membrane protein; phosphoprotein
F:15-94/Domain: GLGF domain homology <GLG1>
F:181-248/Domain: GLGF domain homology <GLG2>
F:416-486/Domain: GLGF domain homology <GLG3>
F:633-782/Domain: guanylate kinase homology <GKI>

Alignment Scores:
Pred. No.: 7,22e-07 Length: 1736
Score: 225.00 Matches: 138
Percent Similarity: 37.88% Conservative: 109
Best Local Similarity: 21.17% Mismatches: 227
Query Match: 5.64% Indels: 179
DB: 2 Gaps: 28

US-09-502-945-5 (1-2162) x A47747 (1-1736)

QY 358 GTGCGTCTGAGCGCTGTCACCCCGAAGCCCTCGGCTGAGTGTGCGGTGGTGGCTGGAG 417
    |||
Db 12 ValThrLeuHisArgAlaProGlyPheGlyPheGlyIleAlaIleSerGlyGlyArgAsp 31
```

```
QY 418 -----TTTGGCTGTGG-----CTCTTCACTCCCACTCATCAAAAGCGGT 459
    |||
Db 32 AsnProHisPheGlnSerGlyGluThrSerIleValIleSerAspValLeuLysGlyGly 51
QY 460 CAGGCAGACAGCGTCCGGCTCAGGTAGGGAGGAGATCGTCGGATCAATGGATATTC 519
    ||||
Db 52 ProAlaGluGly---GlnLeuGlnGluAsnAspArgValAlaMetValAsnGlyValSer 70
QY 520 ATCTCTCTCTGTACCATGAGGAGGTCACTCAACCTCATTCGA---ACCAAGAAACTGTG 576
    |||
Db 71 MetAspAsnValGluHisAlaPheAlaValGlnGlnLeuArgLysSerGlyLysAsnAla 90
QY 577 TCCATCAAGTGTAGACACATCGGC-----CTGATCCCGCTGAAAAGCTCTCTGTATGAG 630
    ||||
Db 91 LysIleThrIleArgArgLysLysValGlnIleProValSerArgProAspProGlu 110
QY 631 CCCTCCTGTCAGTATGTCAGTTCGTGTCGGAATCTCGGGGCTGCGAGGAGC 690
    ||||
Db 111 ProValSerAspAsnGluGluAspSerTyrAspGluGluIleHisAspProArgSerGly 130
QY 691 CTGGGCTCCCTCGAATCGGAAACACAGGAG----- 723
Db 131 ArgSerGlyValValAsnArgArgSerGluLysIleTrpProArgAspArgSerAlaSer 150
QY 723 ----- 723
Db 151 ArgGluArgSerLeuSerProArgSerAspArgSerValAlaSerSerGlnProAla 170
QY 724 AGAAGTCTTCACTACACCTGGTGGTCCCGA-----GGCCTGGCTCGACATT 774
    |||
Db 171 LysProThrLysValThrLeuValLysSerArgLysAsnGluGluTyrGlyLeuArgLeu 190
QY 775 TCAGGCGGCCCTCCAGAACGCTGTCATCATCAGCCATGTGAAACCTGCTCCTG 834
    ||||
Db 191 AlaSer-----HisIlePheValLysGluIleSerGlnAspSerLeu 204
QY 835 TCTGCTGAGGTGGGA---TTTGAGATAGGGGACAGATTCTCGAAGTCAATGGCGTCGAC 891
    ||||
Db 205 AlaAlaArgAspGlyAsnIleGlnGluGlyAspValValLeuLysIleAsnGlyThrVal 224
QY 892 TTCTCTAACTTGGATCACAGGAGGCTGTAATGTGCTGAAATATAGCCGCTGACC 951
    ||||
Db 225 ThrGluAsnMetSerLeuThrAspAlaLysThrLeuIleGluArgSerLysGlyLysLeu 244
QY 952 ATCTCCATTTAGCTGCGAGCTGGCGGAGCTGTTC-----ATGACAGAC 996
    |||
Db 245 LysMetValValGlnArgAspGluArgAlaThrLeuLeuAsnValProAspLeuSerAsp 264
QY 997 CGGAGCGGCTGCGAGAGCGGCGCAGCTGTCAGCTGCGCGCAGGAGCTTCTCATGCAG 1056
    |||
Db 265 SerIleHisSerAlaAsnAlaSerGluArgAsp----- 275
QY 1057 AAGCGCTGGGATGGAGTCCACACAGATCTCCAGGAGCAGGAGATGGAGCGGCAA 1116
    |||
Db 276 -----AspIleSerGluIleGlnSerLeuAlaSerAsp-HisSerGlyLar 290
QY 1117 AGGAAAGAAATTCGCCAGAGCAGCAGAGGAAATAGAGATACCGGAGGAGATG 1176
    |||
Db 290 gSerHisAspArgProProArgArgSerArgSerArgSerProAsp----- 305
QY 1177 GAACAGATTGTAGAGGAGGAGAGAGATTNAGAAGCAATGGGGAAGAGACTGGGGCTCA 1236
    |||
Db 306 -----GlnArgSerGluProSerAspHisSerArgHisSerProGI 319
QY 1237 AAGGAACAGCTACTCTTCCTA-----AAACCATCACT 1269
    |||
Db 319 nGlnProSerAsnGlySerLeuArgSerArgAspGluArgIleSerLysProGlyAl 339
QY 1270 GCTGAGGTACACC-----AGTACCCCTTCGCAAGCCAAAGTATT 1311
    |||
Db 339 aVal-SerThrProValLysHisAlaAspHisThrProLysThrValGluGluValT 359
QY 1312 CCGGAAATATGAGGAAGGCTTTTGACCCCTACTCTATGTTTACCCCGC----- 1358
```



```

Db 359 hrValGluArgAsnGluLysGlnThrProSerLeuProGluProLysProValTyrAlaG 379
QY 1359 -----AGCAGATCATGGG----- 1372
Db 379 lnValGlyAsnGlnMetTrpIleTyrLeuSerValHisLeuMetValSerTyrLeuIleG 399
QY 1373 -----AAGATGTCGGGCTCTACGCATCAAGAAGAGG 1406
Db 399 lnLeuMetLysMetGlyPheLeuArgProSerMetLysLeuValLysPheArgLysGlyA 419
QY 1407 GATCCTTAGACCTGGCCCTGGAAGCGGTGTGACTCCCCCATTTGGGAGGTGGTCGTTT 1466
Db 419 sPSerValGlyLeuArgLeuAlaGlyLysAsp-----ValGly-----IlePheValA 436
QY 1467 CTGCTGTGTATGAGGGGAGCTGTGAGCGCATGTGTGAAGGAGGACGAGAGA 1526
Db 436 laglyValLeuGluAspSerProAlaAlaLysGlu--GlyLeuGluGluLysAspGlnI 455
QY 1527 TCATGCCAATCAACGGCAGATTGTGACAGACTACACCTGCTGAGCTGACGCTGCC 1586
Db 455 leLeuArgValAsnAsnValAspPheThrAsnIleIleArgGluGluAlaValLeuPheL 475
QY 1587 TCAGAAAGCCCTGGAATCAGGCGGGGACTGTGATCGACCTTGTGTGCTGCCCTCC 1646
Db 475 eu-----LeuAspLeu-----p 479
QY 1647 CAAGGAGTATGACGATGAGCTGACCTTCTTGTGAAGTCCAAAGGGAACCAAAATTC 1706
Db 479 roLys-----GlyGluGluValThrIleLeuAlaGlnLysLys----- 492
QY 1707 AGCCCTTAGGAACAAGTGTGCTCGGCCCCCTCTCGTGAACACAAAGCTCGGACCGCC 1766
Db 493 -----AspValTyrArgIleValGluSerAspValGlyAspSerp 507
QY 1767 TTGAG---AGAGGCCACATGACACACACAGATGCATCTTGGACCTGATCATCATC 1823
Db 507 heTyrIleArgThrHisPheGluTyrGluLysGluSerProTyrGly----- 522
QY 1824 CCAGGAATCTCAAACTCCTTTGGCCCTGAACGAGGCGCAGATAAGAACAGCTCGGACC 1883
Db 523 -----LeuSerPheAsnLysGlyGluValPheArgAlaValAspT 536
QY 1884 ACTTTTTCAGAGCAATGTGGAGAAAGGAGGAGCAGCGGCTTTGGGAGAAGATCTCA 1943
Db 536 hrLeuTyrAsnGlyLysLeu---GlySerTrpLeuAlaIleArgIleGlyLysAsnHisL 555
QY 1944 AGGATCCAGACTCTCATCTCTCT-----CTGGCCCGAGTGA 1982
Db 555 ysGluValGluArgGlyIleIleProAsnLysAsnArgAlaGluGlnLeuAlaSerValG 575

```

RESULT 14

Tl0811

channel associated protein of synapse 2 - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000

C:Accession: Tl0811

R:Irle, M.; Hata, Y.; Takai, Y.

submitted to the EMBL Data Library, April 1996

A:Description: Cloning of new isoforms of PSD-95/SAP90 related genes.

A:Reference number: Z17166

A:Accession: Tl0811

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-852 <IRI>

A:Cross-references: EMBL:U53368; NID:g1517939; PID:g1517940

C:Genetics:

A:Note: PSD-95/SAP90-related gene 2 (chapsyn isoform2)

C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase hom

F:198-276/Domain: GLGF domain homology <GLG>
 F:543-601/Domain: SH3 homology <SH3>
 F:663-840/Domain: guanylate kinase homology <GKI>

Alignment Scores:

Pred. No.:	1.51e-06	Length:	852
Score:	218.50	Matches:	120
Percent Similarity:	39.50%	Conservative:	85
Best Local Similarity:	23.12%	Mismatches:	188
Query Match:	5.48%	Indels:	126
DB:	2	Gaps:	24

US-09-502-945-5 (1-2162) x Tl0811 (1-852)

```

QY 307 CAGGTGGAATATGATCAGCTGACCCCGCGCTCCAGGAAGCTGAAGGAGGTGGCTGTG 366
Db 91 GluIleGluTyrGlu-----PheGluGluIleThrLeu 101
QY 367 GACCGTCTGCACCCCGAAGGCTCGGCTGAGTGTGCTGTGGCTGGCTGGAG----- 417
Db 102 GluArgGlyAsn---SerGlyLeuGlyPheSerIleAlaGlyGlyThrAspAsnProHis 120
QY 418 -----TTTGGCTGTGGCTCTTCATCTCCACCTCATCAAAAGCGGTGACGACAGC 471
Db 121 IleGlyAspAspProGlyIlePheIleThrLysIleIleProGlyGlyAlaAlaAlaGlu 140
QY 472 GTCGG---CTCCAGGTAGGAGCAGATCGTCGGATCAATGATATTCATCTCTCTCC 528
Db 141 AspGlyArgLeuArgValAsnAspCysIleLeuArgValAsnGluValAspValSerGlu 160
QY 529 TGTACCCATGAGGAGGTCAATCAAGCTCATTCGA---ACCAAGAAAACTGTCTCCATCAA 585
Db 161 ValSerHisSerLysAlaValGluAlaLeuLysGluAlaGlySerIleValArgLeuTyr 180
QY 586 GTGAGACACATCGGCTGTATCCCGTGAAAGCTCTCTGTATGAGCCCTCCTCTGGCAG 645
Db 181 ValArgArg----- 183
QY 646 TATGTGGATCAGTTGTGTGCGAATCTGGGGCGCTGCGAGCGACCTGGGCTCCCTTGA 705
Db 183 ----- 183
QY 706 AATCGGAAACAAAGAGAGAGGTCTTTCATCAGCTGTGTA---GGCTCCCGAGGCTT 762
Db 184 ArgArgProIleLeuGluThrValGluIleLysLeuPheLysGlyProLysGlyLeu 203
QY 763 GGCTGACGATTCACGCGGCCCATCCAGAAAG-----CTGGC-----ATCTTT 807
Db 204 GlyPheSerIleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyr 223
QY 808 ATCAGCCATGTGAACCTGGCTCCCTGCTGCTGAGTGAGTGGGA---TTGGAGATAGGGAC 864
Db 224 ValThrLysIleIleAspGlyGlyAlaAlaGlnLysaspGlyArgLeuGlnValGlyAsp 243
QY 865 CAGATTGTCGAAGTCAATCGCTCGACTTCTCTAACCTGGATCACAAGGAGGTGTAAT 924
Db 244 ArgLeuLeuMetValAsnAsnTyrSerLeuGluGluValThrHisGluGluAlaAla 263
QY 925 GTGCTGAAAATAGCCCGACGCTGACCATCTCCATCTCCATCTGAGCTGCGCGGAGCTG 984
Db 264 IleLeuLysAsnThr---SerAspValValTyrLeuLysValGlyLysProThrIle 282
QY 985 TTATGACACACCGGGGCGGCTGGCAGAGCGGCGGCGGCGAGCTGAGCTGACGCGCAGGAG 1044
Db 283 TyrMetThrAspProTyrGlyProProAspIle-----ThrHisSerTyrSer 298
QY 1045 CTCTCATGCAGAAGCGCTGGCGATGGAGTCCCAACAAGATCTCCAGGAGCAGCAGGAG 1104
Db 299 ProProMetGluAsnHisLeuLeuSerGlyAsnGlnGlyThrLeuGlyTyrLysThrSer 318
QY 1105 ATG-----GAGCGGCAAGAGGAAAAAAGAAATTCGCCAGAGGAGCAGAGGAGAAAT 1155
Db 319 LeuProProIleSerProGlyArgTyrSerProIleProLysHisMetLeuValGluAsp 338

```



```

Db 364 rSerProValGluCysAspLysSerPheLeuLeuSerAlaProTyrSerHisTyrHisLe 384
QY 1327 AGCCTTTGACCCCTACTATGTTACCCAGAG----- 1360
Db 384 uGlyLeuLeuProAspSerGluMetThrSerHisSerGlnHisSerThrAlaThrArgGl 404
QY 1361 -----CAGATCATGGGAAGGATGTCGGCTCCTACG 1392
Db 404 nProSerMetThrLeuGlnArgAlaValSerLeuGluGlyGluProArgLysValValLe 424
QY 1393 CATCAGAGAGGGATCCTTAGACCTGGCCCTGGAGCGGTGTGGACTCCCCCATGG 1452
Db 424 uHisLysGlySerThrGlyLeuGlyPheAsnIleValGlyGlyGluAspGlyGlu----- 442
QY 1453 GAAGGTGGTCTTCTGCTGTATGAGGGGAGCTGCTGAGCGGCATGGTGGCATGCT 1512
Db 443 -GlyIlePheValSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGluLeuGl 462
QY 1513 GAAAGGGACGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCCCTGGCTGA 1572
Db 462 nArgGlyAspGlnIleLeuSerValAsnGlyIleAspLeuArgGlyAlaSerHisGluGl 482
QY 1573 GGCTGACGCTGCCCTGCAGAGCCCTGGNAATCAGGCGGGGAGCTGGATCGACCTTGTGCT 1632
Db 482 nAlaAlaAlaAlaLeu-----LysGlyAlaGlyGlnThrValThrIleIl 497
QY 1633 TGCCGTCTGCCCCCAAGGAGTATGACGATGACGTGACCTTCTTCTGCTGAAGTCCAAAG 1692
Db 497 eAlaGlnTyrGlnProGluAspTyr-----AlaArgph 508
QY 1693 GGGAAACCAATTCACGCTTAGGAAACAGTACGCTCGGCCCCACCTCGTGAAC 1747
Db 508 eGluAlaLysIleHis-----AspLeuArgGluGlnMetMetAsn 521

```

Search completed: March 21, 2003, 13:06:25
Job time : 72.5119 secs

Result	No.	Score	Query	Match	Length	DB	ID	Description
1	2648.5	66.4	552	1	A175_HUMAN		Q9Y6N9	homo sapien
2	272	6.8	905	1	Z03_MOUSE		Q9QXV1	mus musculus
3	271	6.8	933	1	Z03_HUMAN		Q95049	homo sapien
4	255.5	6.4	898	1	Z03_CANFA		Q62683	canis famil
5	249.5	6.3	849	1	DLG3_MOUSE		P70175	mus musculus
6	249.5	6.3	849	1	DLG3_RAT		Q62936	rattus norv
7	236	5.9	1745	1	Z01_MOUSE		F39447	mus musculus
8	234.5	5.9	767	1	DLG4_HUMAN		P78352	homo sapien
9	232.5	5.8	724	1	DLG4_RAT		F31016	rattus norv
10	232.5	5.8	817	1	DLG3_HUMAN		Q92796	homo sapien
11	232	5.8	904	1	DLG1_HUMAN		Q12959	homo sapien
12	229.5	5.8	724	1	DLG4_MOUSE		Q62108	mus musculus
13	226	5.7	911	1	DLG1_RAT		Q62696	rattus norv
14	225	5.6	1736	1	Z01_HUMAN		Q07157	homo sapien
15	215.5	5.4	852	1	DLG2_RAT		Q63622	rattus norv
16	213.5	5.4	870	1	DLG2_HUMAN		Q15700	homo sapien
17	204.5	5.1	2485	1	P7ND_HUMAN		Q12923	homo sapien
18	195.5	4.9	2161	1	SHK1_HUMAN		Q9Y566	homo sapien

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF039700; AAC18049.1; ALT_FRAME.
 DR EMBL; AF039699; AAC18048.1; -
 DR EMBL; AB006955; BAA81739.1; -
 DR EMBL; AB018687; BAA81740.1; -
 DR HSSP; P29476; 1QAV.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 6.
 DR SMART; SM00228; PDZ; 3.
 DR PROSITE; PS01016; PDZ; 3.
 KW Antigen; Repeat; Alternative splicing; Polymorphism.
 FT DOMAIN 87 169
 FT DOMAIN 211 293
 FT DOMAIN 452 537
 FT VARSPPLIC 1 31 MISSING (IN ISOFORM 2).
 FT VARSPPLIC 404 552 MISSING (IN ISOFORM 3).
 FT VARIANT 519 519 /FTID=VAR_012320.
 FT CONFLICT 103 103 R -> S (IN REF. 2; BAA81739).
 FT CONFLICT 280 280 N -> S (IN REF. 2).
 FT CONFLICT 305 305 A -> T (IN REF. 2; BAA81739).
 SQ SEQUENCE 552 AA; 62223 MW; 7B405B02CACE8241 CRC64;

Alignment Scores:

Pred. No.: 5,73e-145 Length: 552
 Score: 2648.50 Matches: 525
 Percent Similarity: 95.2% Conservative: 1
 Best Local Similarity: 95.11% Mismatches: 0
 Query Match: 66.40% Indels: 26
 DB: 1 Gaps: 1

US-09-502-945-5 (1-2162) x A175_HUMAN (1-552)

QY 97 ATGGACCGAAAGTGCCCGCAGAAATCCGGCATAGGTGGATTTCTGTGATGCAAAATGAT 156
 Db 1 MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp 20
 QY 157 GCAGAGAAGGACTATCTCTATGATGTCGCGAATGTACACACAGACCATGACGTGGCC 216
 Db 21 AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla 40
 QY 217 GTGCTCGTGGGAGACCTGAAGCTGTGTCATCAATGAACCCAGCCGCTGCCTCTGTTGAT 276
 Db 41 ValLeuValGlyAspLeuLysLeuValIleAsnGluProSerArgLeuProLeuPheAsp 60
 QY 277 GCCATTCCGCCCGCTGATCCACTGAGCAGCAGGTCGGATATGATCAGCTGACCCCGG 336
 Db 61 AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg 80
 QY 337 CGCTCCAGGAAGCTGAAGAGGTGCTGTGGACCGTCTGCACCCGGAAGCCCTCGGCCCTG 396
 Db 81 ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu 100
 QY 397 AGTGTGCGTGGTGGCGCTTGGGCTCTTCCACCCGATCCATCAAGGC 456
 Db 101 SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly 120
 QY 457 GGTGAGCAGACAGCTCGGGCTCAGGTAGGGGAGCAGATCGTCGGATCAATGATAT 516
 Db 121 GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr 140
 QY 517 TCCATCTCTCTCCCTACCATGAGGAGTCATCAACCTCATTCGACCAAGAAACTG 576
 Db 141 SerIleSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal 160
 QY 577 TCCATCAAGTGAACACATCGCCCTGATCCCGTGAAAGACTCTCTCTGATGAGCCCTC 636

Db 161 SerIleLysValArgHisIleGlyLeuIleProValLysSerProAspGluProLeu 180
 QY 637 ACTTGGCAGTATGGGATCAGTTTGTCTCGAAATCTCGGGCGCTGCGAGCAGCCTGGGC 696
 Db 181 ThrIrpGlnTyrValAspGlnPheValSerGluSerGlyValArgGlySerLeuGly 200
 QY 697 TCCCTCGAAATCGGAAAAACAAGGAGAAGAGTCTTCATCAGCCTGTGGTAGGCTCCCGA 756
 Db 201 SerProGlyAsnArgGluAsnLysGluLysValPheIleSerLeuValGlySerArg 220
 QY 757 GGCCTTGGCTGCACATTTCCAGCGGCCCATCCAGAGCCTGGCATCTTATCAGCCAT 816
 Db 221 GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheIleSerHis 240
 QY 817 GTGAAACCTGGCTCCCTGTCTGTGAGTGGGATTGAGATAGGGAGCCAGGATGTGCAA 876
 Db 241 ValLysProGlySerLeuSerAlaGluValGlyLeuGluIleGlyAspGlnIleValGlu 260
 QY 877 GTCAATGGCTGCTGCTTCTTCACTGATCACAAAGGAGGCTGTAAATGTCTGCTGAAAAT 936
 Db 261 ValAsnGlyValAspPheSerAsnLeuAspHisLysGluAlaValAsnValLeuLysAsn 280
 QY 937 AGCCGCGACCTGACCATCTCCATTGTAGCTGCAGCTGCCGGGAGCTGTTTCATCAGAC 996
 Db 281 SerArgSerLeuThrIleSerIleValAlaAlaAlaGlyArgGluLeuPheMetThrAsp 300
 QY 997 CGGAGGCGCTGGCAGAGCGCGCAGCTGAGCTGCAGCGGCGAGGAGCTTCTCATGCAG 1056
 Db 301 ArgGluArgLeuAlaGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuMetGln 320
 QY 1057 AAGCGGCTGGCGATGGAGTCCCAACAAGATCTCCAGGAGCAGCAGGAGATGGAGCGCAA 1116
 Db 321 LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGluGlnGluMetGluArgGln 340
 QY 1117 AGGAGAAAGAAATGGCCAGAGCAGCAGAGAAATAGAGATATCCGGAAGAGATG 1176
 Db 341 ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet 360
 QY 1177 GAACAGATTGTAGAGGAGGAAGAAGTTTAAAGAACAAATGGGAAGAGCTGGGGCTCA 1236
 Db 361 GluGlnIleValGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer 380
 QY 1237 AAGGAACAGCTACTCTTGGCTTAAACCATCATCTGCTGAGGTACACCCAGTACCCTTGGC 1296
 Db 381 LysGluGlnLeuLeuLeuProLysThrIleThrAlaGluValHisProValProLeuArg 400
 QY 1297 AAGCCA----- 1302
 Db 401 LysProLysTyrAspGlnGlyValGluProGluLeuGluProAlaAspLeuAspGly 420
 QY 1303 -----AAGTGATTTCGGAAATATAGGAAGGCTTTGACCCCTA 1341
 Db 421 GlyThrGluGluGlnGlyGluGln-AspPheArgLysTyrGluGluGlyPheAspProTy 440
 QY 1342 CTCTATGTTCACCCAGCAGCAGATCATCGGGAAGGATCTCCGGCTCTTACCGATCAAGAA 1401
 Db 440 rSerMetPheThrProGluGlnIleMetGlyLysAspValArgLeuLeuArgIleLysLys 460
 QY 1402 GGAGGATCTTAGACCTGGCCCTGGGAAGCGGTGTGACCTCCCGCATTTGGGAGGTGCT 1461
 Db 460 sGluGlySerLeuAspLeuAlaLeuGluGlyGlyValAspSerProIleGlyLysVal 480
 QY 1462 CGTTTCTCTGTGTATGATGAGCGGGAGCTGCTGAGCGGATGTGTGCAATGTGAAGGGA 1521
 Db 480 lValSerAlaValTyrGluArgGlyAlaAlaGluArgHisGlyIleValLysGlyAs 500
 QY 1522 CGAGATCATGGCAATCAACGGCAGCATTTGTGACAGACTACACCTGGCTGAGGCTGACGC 1581
 Db 500 pGluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaAspAl 520
 QY 1582 TGCCCTCAGAAAGCGCTGGAATCAGGCGGGAGCTGGATCGACCTGTGGTTCGCTGCTG 1641

Db 520 aAlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValValalaValcy 540

QY 1642 CCCCCCAAGGAGTATGAGTACGAGTACGCTGC 1675

Db 540 sProProlysGluTyrAspAspGluLeuThrPhe 551

RESULT 2

Z03_MOUSE

ID Z03_MOUSE STANDARD; PRT; 905 AA.

AC Q90XV1.

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona occludens 3 protein) (Tight junction protein 3).

GN TJP3 OR Z03.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALE/C;

RX MEDLINE=20069797; PubMed=10601346;

RA Itoh M., Furuse M., Morita K., Kubota K., Saitou M., Tsukita S.

RT "Direct binding of three tight junction-associated MAGUKs, ZO-1, ZO-2, and ZO-3, with the COOH termini of claudins.";

RL J. Cell Biol. 147:1351-1363(1999).

CC -|- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1.

CC -|- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.

CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC -|- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AF157006; AAF24175.1; -.

DR HSP; P31016; IBEF.

DR MGD; MGI:1351650; Tjp3.

DR InterPro; IPR000619; Guanylate_kin.

DR InterPro; IPR001478; PDZ.

DR InterPro; IPR001452; SH3.

DR Pfam; PF00595; PDZ; 3.

DR Pfam; PF00625; Guanylate_kin; 1.

DR SMART; SM00072; GuK; 1.

DR SMART; SM00228; PDZ; 3.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS00856; GUANYLATE_KINASE_1; FALSE_NEG.

DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.

DR PROSITE; PS00002; SH3; FALSE_NEG.

DR PROSITE; PS0106; PDZ; 3.

KW Tight junction; SH3 domain; Repeat; Membrane.

FT DOMAIN 11 93

FT PDZ 1.

FT DOMAIN 187 264

FT PDZ 2.

FT DOMAIN 368 434

FT PDZ 3.

FT DOMAIN 467 540

FT SH3.

FT DOMAIN 654 754

GUANYLATE_KINASE.

SQ SEQUENCE 905 AA; 99324 MW; B787BA1592661FEE CRC64;

Alignment Scores:

Pred. No.: 2,06e-08 Length: 905

Score: 272.00 Matches: 144

Percent Similarity: 38.45% Conservative: 89

Best Local Similarity: 23.76% Mismatches: 203

Query Match: 6.82% Indels: 171

DB: 1 Gaps: 30

US-09-502-945-5 (1-2162) x Z03_MOUSE (1-905)

QY 395 GGCCTCGGCTGAGTGTGCGTGGCTGGAGTTTGGCTGTGGG---CTCTTCATCTCC 441

Db 21 GlyPheGlyIleAlaValSerGlyGlyHisAspArgAlaSerGlySerValValSer 40

QY 442 CACCTCATCAAGGCGGTTCAGGCAGACAGCTCGGGCTCCAGGTAGGGAGGAGATGCTC 501

Db 41 AspValValProGlySerProAlaGluGly---ArgLeuArgThrGlyAspHisIleVal 59

QY 502 CGGATCAATGATATTCATCTCCTGTACCATGAGGAGGTCAATCAACCTCATTCGA 561

Db 60 MetValAsnGlyValSerValGluAsnValThrSerAlaPheAlaIleGlnIleuLys 79

QY 562 ACC---AAGAAACTGTGTCATCAAAAGTG-----AGACACATGCCTGATCCCC 609

Db 80 ThrCysThrLysThrAlaAsnValThrValLysArgProArgValGlnLeuProAla 99

QY 610 GTGAAAAGCTCTCTCTGATGAGCCCTCACCTGGCAGTATGTGGATCAGTTTGTCTCGAA 669

Db 100 ThrLysAlaSerProAlaSer-----GlyHisGlnLeuSerAspGlnGluAlaAsp 117

QY 670 TCTGGG----- 675

Db 118 HisGlyArgGlyTyrGluGlyAspSerSerGlySerGlyArgSerTrpGlyGluArg 137

QY 676 -----GGCTGCGAGGAGGAGCTGGCTCCCTCGGAAATCGGGAAC 717

Db 138 SerArgArgSerArgAlaGlyArgGlyArgValGlySerHisGlyArgArgSer 157

QY 718 -----AAG 720

Db 158 GlyGlySerGluAlaAsnGlyLeuAspLeuValSerGlyTyrLysArgLeuProLys 177

QY 721 GAG-----AAGAGGTCTTCATCAGCCTGTGTAGCTCCCGAGGC 759

Db 178 GlnAspValLeuMetArgProLeuLysSerValLeuValLysArgArgAsnSerGluGlu 197

QY 760 CTGGCTGCGAGCTTTCACGGCCCATCCAGAGCCTGCATCTTATTCAGCCATGTG 819

Db 198 PheGlyValLysLeuGlySer-----GlnIlePheIleLysHisIle 211

QY 820 AAACCTGGCTCCCTGTCTGTCTGCT---GAGGTGGATTTGGAGATAGGCACCATGTCGAA 876

Db 212 ThrGluSerGlyLeuAlaAlaArgAsnHisGlyLeuGlnGluGlyAspLeuIleLeuGln 231

QY 877 GTCATGCGCTCGACTTCTCTAACCTGGATCACAAGGAGGCTGTAATGTGCTGAAAAAT 936

Db 232 IleAsnGlyValSerSerAlaAsnLeuSerLeuSerAspThrArgArgLeuIleGluLys 251

QY 937 AGCCGC---AGCCTGACCATCTCCATTTGATGTCAGCTGGCGGGAGCTGTTTCATGACA 993

Db 252 SerGluGlyGluLeuThrLeuLeuValLeuArgAspSerGlyGlnPheLeuVal---AsnI 271

QY 994 GACCGGAGCGGTGGCGAGGCGCGCGCTGAGCTGCGAGCGGAGGAGCTTCTCATG 1053

Db 271 leProProAla-----ValSerAspSerAspSerSerLeu 283

QY 1054 CAGAAGCGCTGGCGAT-----GGAGTCCAACAAGATCTCCAGGAGCAGCAGGAG 1104

Db 283 etGluAspIleSerAspLeuThrSerGluLeuSerGlnAlaProProSerHisValProp 303

QY 1105 ATGAGCGGCAAGGAGAGAAAGAAATGCCCAAGGAGCAGCAGGAGGAGGAGGAGATAC 1164

Db 303 roProProLeuLysGlyGlnArgSer-----p 312

QY 1165 CGGAAGGAGATGAACAGATTTGTAG-----AGGAGGAGGAGGAGGAGGAGGAG 1202

Db 312 roGluAspSerGlnThrAspSerProValGluThrProGlnProArgArgGluArgS 332

QY 1203 GTTTAAGAACCAATGGGAAGAGACTGGGCTCAAAGGAGCAGCTACTCTTGCCTAAAC 1262

Db 332 erValAsnSer-----ArgAlaIleAlaGluProGluSer-----p 344


```
Db 100 ThrLysAlaSerProSerSerProGlyArgGlnAspSerAspGluAspAspGlyProGln 119
QY 637 ACTTGGCAGTATGTGATCAG-----TTTGTGTCGGAATCNGG 675
Db 120 ArgValGluGluValaspGlnGlyArgGlyTyraSpGlyaspSerSerGlySerGly 139
QY 676 -----GGCGTGCAGGAGCAGCCTGGGCTCC 699
Db 140 ArgSerTrpAspGluArgSerArgProArgProGlyArgGlyArgAlaGlySer 159
QY 700 CCTGGAATCGGAAACAAGGAGAAGGCTTTCATCAGCCTGGTAGCTCCGAGGC 759
Db 160 HisGlyArgSerProGlyGly-----GlySerGluAla 171
QY 760 CTTGGTGCAGCATTTCCAGCGC-----CCCATCCAG-----AAG 795
Db 172 AsnGlyLeuAlaLeuValSerGlyPheLysArgLeuProArgGlnaspValGlnMetLys 191
QY 796 CCH----- 798
Db 192 ProValLysSerValLeuValLysArgArgAspSerGluGluPheGlyValLysLeuGly 211
QY 799 ---GGCATCTTTATCAGCCATGTGAACCTGGCTCCTGCTGCTCT---GAGGTGGATTC 852
Db 212 SerGlnPheIleLeuLysHisIleThrAspSerGlyLeuAlaAlaArgHisArgGlyLeu 231
QY 853 GAGATAGGGAGCAGATTCGGAAGTCAATGGCGTGGACTTCTTAACCTGGATCACAAG 912
Db 232 GlnGluGlyAspLeuIleAsnGlyValSerSerGlnAsnLeuSerLeuAsn 251
QY 913 GAGCGTGTAAATGCTGTAATAATACCGGAGCGCTGACCATCCTCCATTCAGTCAGCT 972
Db 252 AspThrArgLeuLeuGlySerGlySerGlyGly---LysLeuSerLeuLeuValLeuArg 270
QY 973 GGCGGCGAGCTGTTCATGACAGACCGGAGCGCGCTGGCAGCGCGCGGAGCTGAGCTG 1032
Db 271 AspArgGlyClnPheLeuValAsnIleProProAlaValSerAspSerAspSerPro 290
QY 1033 CAGCGCGAGGAGCTTCTCATGAGAGAGCGGCTGGCGATGAGTCCACAAGATCCTCCAG 1092
Db 291 LeuGluGluGlyValThrMetAlaaspGluMetSerSerProProAlaaspSerAsp 310
QY 1093 GAGCAGCAGGATGGAGCG----- 1113
Db 311 LeuAlaSerGluLeuSerGlnAlaProProSerHisIleProProProArgHisAla 330
QY 1114 CAAGGAGAAAGAAATTCGCCAG---ANGGAGCAGAGAGAAATCAGAGATACCGGAG 1170
Db 331 GlnArgSerProGluAlaSerGlnThrAspSerProValGluSerProArgLeuArgArg 350
QY 1171 GAGATG-----GAACAGATTGTAGAGGAGGAGAGAGTTTAAAGAACATGGGAA 1221
Db 351 GluSerSerValaspSerArgThrIleSerGluProAspGluGlnArgSerGlu----- 368
QY 1222 GAAGACTGGGCTCAAGGAGACAGCTACTCTTCCTTAAACCATCATCCTGCTGAGTAC 1281
Db 369 -----LeuProArgGluSerSerTyraSpIleTy 378
QY 1282 CCAGTACCCCTTCGCAAGCAAGTCATTCCTCGGAAATATGAGGAGCGTTTGCACCCCTA 1341
Db 379 ArgVal-Pro-----SerSerGlnSer-----MetGluAspArgGlyTyraSerPro-- 393
QY 1342 CTCATGTTACCCAGCAGCATCATGGGAGAGATGTCGGCTCCCTACGATCAAGAA 1401
Db 394 -----AspThrArgValValArgPheLeuLys 402
QY 1402 GGAGGATCCTTAGACCTGGCCCTGGAAGCGGTGTGGACTCCCTCCCATTTGGGAAGGTGT 1461
Db 402 sclyLysSerIleGlyLeuArgLeuAlaGlyLysAsnAsp-----ValGly---IlePh 419
QY 1462 CTTTCTGCTGTATGACGGGGAGCTGCTGAGCGGATGTGTGATGTGAAGAGGGA 1521
Db 419 eValSerGlyVal---GlnAlaGlySerProAlaAspGlyGlnGlyIleGlnGluGlyAs 438
```

```
QY 1522 CGAGATCATCGGCAATCAACGGCAAGATTGTGACAGACTACACCTGGCTGAGGCTGACGC 1581
Db 438 pGlnIleLeuGlnValAsnAspValProPheGlnAsnLeuThrArgGluGluAla----- 456
QY 1582 TGCCTTCGCAAGAGCCCTGGAATCAGGCGGGGACTGATCGACCTTGTGGTTGCCCTCTCG 1641
Db 457 -----ValGlnPheLeuLeuGlyLeuPr 464
QY 1642 CCCCCCAAGGAGTATGAC-----GATGAGCTGACCTTCTTGTCT 1680
Db 464 oProGlyGluGluMetGluLeuValThrGlnArgLysGlnAspIlePheTrpLysMetVa 484
QY 1681 GAAGTCCAAAGGGGAAACCAATTCACCGGTAGGAACAGTGAAGTCCGGCCGCCACCT 1740
Db 484 IginSerArgValGlyAspSerPheTyrlleArgThrHisPheGluLeuGluPro----- 502
QY 1741 COTGAACAAAGCCTCGGACCAAGCCTT-----GAGAGAGGCCACATCACACACACC-- 1792
Db 503 -----SerProSerGlyLeuGlyPheThrArgGlyAspValPheHisValle 519
QY 1793 ---AGATGGCATCTTGGGACCTGAATCTATCACCAGGAATCTCAAACTCCCTTTGGC 1848
Db 519 uAspThrLeuHisPro----- 524
QY 1849 CCGTGAACAGGCGGAGATAAGGAACAGCTCGGGCCACTTTTGAAGGCCAATGTGGAGG 1908
Db 525 ---GlyProGlyGlnSerHisAlaArgGlyGlyHisTrpLeu----- 537
QY 1909 AAGGAGCAGCAGCCGCTTGGGAGAGATCTCAAGATCCAGACTCTCATTCCTTTCC 1968
Db 538 -----AlaValArgMetGlyArgAspLeuArgGluGlnGluArgGlyIlelePr 554
QY 1969 T 1969
Db 554 o 554
RESULT 4
203_CANFA
ID ZO3_CANFA STANDARD; PRT; 898 AA.
AC O62683;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tight junction protein ZO-3 (Zonula occludens 3 protein) (Zona
DE occludens 3 protein) (Tight junction protein 3).
GN TJP3 OR ZO3.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=98198478; PubMed=9531559;
RA Haskins J., Gu L., Wittchen E.S., Hubbard J., Stevenson B.R.;
RT "ZO-3, a novel member of the MAGUK protein family found at the tight
RT junction, interacts with ZO-1 and occludin.";
RL J. Cell Biol. 141:199-208(1998).
CC -|- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-1.
CC -|- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AF023617; AAC39177.1; -
```


QY 209 ACGTGCCGCTGCTCGTGGGAGACCTGAAGCTGGTTCATCAATGAACCCAGCCGCTGCTC 268
 Db 117 -----TyrGluGlnAlaSerProAlaPro---Leu 125
 QY 269 TCTTTGATGCCATTCGGCGGTGATCCCACTGAACACACAGGTGGAATATGATCAGCTGA 328
 Db 126 LeuValAsnPro-GluAlaLeuGluProSerLeuSerValAsnGly-SerAspGlyMetP 145
 QY 329 CCCCCCGGCTCCAGGAGGCTGAAGAGGTCGCTGCTGAGCCGCTGACCCCGGAGGCC 388
 Db 145 he-----LysTyrGluGluIleValLeuGluArgGlyAsn---SerGlyL 159
 QY 389 TCGGCTGAGTGTGGTGGCTGGAG-----TTTGGCTGTGGCTC 433
 Db 159 euGlyPheSerIleAlaGlyIleAspAsnProHisValProAspAspProGlyIleP 179
 QY 434 TCATCTCCACCTCATCAAGCGGTGAGGAGAGCGGTGGG---CTCAGGTAGGG 490
 Db 179 heIleThrLysIleIleProGlyAlaAlaAlaMetAspGlyArgLeuGlyValAsnA 199
 QY 491 ACGAGATCGTCCGATCAATGATATCCATCTCTCTCTGCTGACCATCAGGAGGTATCA 550
 Db 199 spCysValLeuArgValAsnGluValAspValSerGluValHisSerArgAlaValG 219
 QY 551 ACCTCATTCGAACCAAGAAACTGTGTCCATCAAGTGAAGACATCGGCTGATCCCG 610
 Db 219 luAlaLeuLysGluAlaGlyProVal-----ValArg-----LeuValV 232
 QY 611 TGAAGATCTCTGATGAGCCCTCACTTGGCGATGATGTGATCAGTTGTGTGCGGAAT 670
 Db 232 alaArgArgGlnProProGluThrIleMetGluValAsn----- 246
 QY 671 CTGGGGCGTGGAGCGACCTGGGCTCCCTCGGAATCGGAAACACAGGAGAGAGG 730
 Db 246 ----- 246
 QY 731 TCTTCATCAGCTGTAGGCTCCGAGCGCTTGGTGGCAGCATTTCCAGCGGCCCATCC 790
 Db 247 -----LeuLeuLysGlyProLysGlyLeuGlyPheSerIleAlaGlyIleGlyA 264
 QY 791 AGAAG-----CCTGGC-----ATCTTATCAGCCATGTGAACCTGGCTCCTGT 835
 Db 264 snGlnHisIleProGlyAspAsnSerIleTyrIleThrLysIleIleGlyAlaA 284
 QY 836 CTGCTGAGGTGGGA---TTGAGATAGGGACAGATTGTGCAAGTCAATGGCGTCACT 892
 Db 284 laGlnLysAspGlyArgLeuGlnIleGlyAspArgLeuLeuAlaValAsnAsnThrAsnL 304
 QY 893 TCTTAACCTGGATCAAGAGGCTGTAATGTGCTGAAATATAGCCGAGCCTGACCA 952
 Db 304 euGlnAspValArgHisGluAlaValAlaSerLeuLysAsnThrSerAspMetVal 324
 QY 953 TCTCCATTGATGAGTGCAGTGGCGGAGCTGTCATGACAGACCGGCGGCTGCGAG 1012
 Db 324 yrLeuLysValAlaLysProGly---SerLeuHisLeuAsnAsp----- 337
 QY 1013 AGCGCGGACGCTGAGCTGCAGCGGAGGAGCTTCTCATGACAGAGCGGTGCGCATGG 1072
 Db 338 -----MetTyrAlaProProAspTyrAlaSerThrPheThrAlaLeuA 352
 QY 1073 AGTCCACAGATCTCTC---AGGAGCAGGAGGATGGAGCGCAAGGAGGAGGAAA 1129
 Db 352 laAspAsnHisIle-SerHisAsnSerSerLeuGlyIleValGluSerLys 371
 QY 1130 TTGCCAGAGGAGCAGAGAAATGAGATACCCGAGGAGATGGAACAGATTGTAG 1189
 Db 372 Val----- 372
 QY 1190 AGGAGGAAGAGAGCTTTAAGAAGCAATGGGAAGAAGACTGGGGCTCAAGAGACAGCTAC 1249
 Db 372 ----- 372

QY 1250 TCTTGCCCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCAAGTGAT 1309
 Db 373 -----ThrTyrProAlaProProGlnValPro 381
 QY 1310 TTCGGAAATATGAGGAAGGCTTTGACCCCTACTCTATTCACCCCGACAGCAGATCATG 1369
 Db 382 ProThrArgTyrSerPro-----IleProArgHisMetLeuAlaGluGluAspPheThr 399
 QY 1370 GGAAGAGATGTCGGCTCTACGATCAAGAAGAGGAGGATCTTTAGACCTGGCCCTGAA 1429
 Db 400 ArgGluProArgLysIleIleLeuHisLysGlySerThrGlyLeuGlyPheAsnIleVal 419
 QY 1430 GGCGGTGAGGACTCCCATTTGGGAAGTGGTCTCTCTGCTGTGTATGAGCGGGAGCT 1489
 Db 420 GlyGlyGluAspGlyGlu-----GlyIlePheValSerPheIleLeuAlaGlyGlyPro 437
 QY 1490 GCTGAGGGGATGTCGTCATTTGAAAGGGGACGAGATCATGGCAATCAACGGCAAGATT 1549
 Db 438 AlaAspLeuSerGlyGluLeuArgArgGlyAspArgIleLeuSerValAsnGlyValAsn 457
 QY 1550 GTGACAGATACACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1609
 Db 458 LeuArgAsnAlaThrHisGluGlnAlaAlaAlaAlaLeuLysArgAla----- 473
 QY 1610 GGGGACTGGATCGACCTTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1669
 Db 474 GlyGlnSerVal-----ThrIleValAlaGlnTyrArgProGluGluTyr----- 488
 QY 1670 ACCTTCTTGTGAAGTCCAAAGGGGAAACCAATTCACGCGTTAGGAAACAGTGAGCTC 1729
 Db 489 -----SerArgPheGluSerLysIleHis-----AspLeu 498
 QY 1730 CGGCCCCACCTCGTGAACACAAAGCCT-----CGGACACAGCCTT 1768
 Db 499 ArgGluGlnMetMetAsnSerSerMetSerGlySerGlySerLeuArgThrSerGlu 518
 QY 1769 GAGAGA 1774
 Db 519 LysArg 520
 RESULT 7
 ZOI_MOUSE STANDARD; PRT; 1745 AA.
 ID ZOI_MOUSE AC P39447;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tight junction protein ZO-1 (Zonula occludens 1 protein) (Zona
 occludens 1 protein) (Tight junction protein 1).
 GN TP1 OR ZO1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=93352986; PubMed=8486731;
 RA Itoh M., Nagafuchi A., Yonemura S., Yasuda-Kitani T., Tsukita S.,
 RA Tsukita S.;
 RT "The 220-kD protein colocalizing with cadherins in non-epithelial
 RT cells is identical to ZO-1, a tight junction-associated protein in
 RT epithelial cells: cDNA cloning and immunoelectron microscopy.";
 RL J. Cell Biol. 121:491-502(1993).
 CC -!- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL
 CC REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY
 CC HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN
 CC MIGHT BE INVOLVED IN STABILIZING JUNCTIONS.
 CC -!- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-3.
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.
 CC MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT
 CC OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.

Db	443	----	ValGly	----	IlePheValAlaGlyValLeuGluAspSerProAlaAlaLysGlu	--	459
Qy	1504	TGGCA	TTGTGA	AGGGGAGAGATCAT	GCATCAATCAACGCGCAAGATTGTGCACAGACTACAC		1563
Db	460	-GlyLeu	GluGluGlyAsp	GlnIleLeuArgValAsnValAspPheThrAsnIle	I	479	
Qy	1564	CTTGCGT	GAGCGTCCCTCGAC	GAAGCCCTGGAAATCAGCGCGGGGACTGCGATCGA		1623	
Db	479	eArgGlu	GluValAlaValLeuPheLeu			489	
Qy	1624	CCITGT	GTGGTCCGCTGCTG	CCCCCAAGGAGTATGACGATGAGCTGACCTCTTCTGCTGAA		1683	
Db	489	pLeu	-----	ProLys	-----GlyGluGluValThrIleLeuAlaG	501	
Qy	1684	GTCCAA	AGGGGAAACAAATTC	ACCGCTTAGAACACAGTGCAGCTCCGCGCCACCTCGT		1743	
Db	501	nLysLys	Lys	-----	AspValTyrArgAlaGluLeuVal	511	
Qy	1744	GAACACA	AAAGCCCTCGGAC	CAGCTTGAG	---AGAGGCCACATGACACACACACATGGCA	1800	
Db	511	lGluSer	AspValGlyAspSerPheTyrIle	ArgThrHisPheGluTyrGluLysGlu	Se	531	
Qy	1801	TCCTTG	GGACCTGAATCTATC	ACCCAGGAATCTCAACTCTTGGCCCTTGAACACAGG		1860	
Db	531	rProTyr	Gly	-----	LeuSerPheAsnLysG	540	
Qy	1861	CCAGATA	AGGAACAGCTCG	CGGCCACTTTT	TGAAGGCCCAATGTGGAGGAAAGGAGCAGC	1920	
Db	540	yGluVal	PheArgValValAspThrLeuTyrAsnGlyLysLeu	--	GlySerTrpLeuAl	559	
Qy	1921	CAGCGT	TGGGAGAGATCTC	ANGATCCACACTCTCATCTCTTCTCT		1969	
Db	559	AlaArgIle	GlyLysAsnHisLysGluValGluArgGlyIle	IleProAsnLysAsnAr		579	
Qy	1970	-----	CTGCCCCAGTGAAT	TGGTCTCTCCCA	-----GCTTTGGGGGAC	2008	
Db	579	gAlaGlu	GlnLeuAlaSerValGlnTyrThrLeuProLysThrAlaGlyGlyAsp			597	
RESULT 8							
ID	DLG4_HUMAN	STANDARD;	PRT;	767	AA.		
AC	P7352; Q92941; Q9UKK8;						
DT	01-NOV-1997 (Rel. 35, Created)						
DT	30-MAY-2000 (Rel. 39, Last sequence update)						
DT	15-JUN-2002 (Rel. 41, Last annotation update)						
DE	Presynaptic density protein 95 (PSD-95) (Discs, large homolog 4)						
DE	(Postsynaptic density-95).						
GN	DLG4 OR PSD95.						
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
OX	NCBI_TaxID=9606;						
RP	[1]						
RC	SEQUENCE FROM N.A.						
RC	TISSUE=Mammary gland;						
RC	MEDLINE=9743282; Pubmed=9286702;						
RX	Stathakis D.G., Hoover K.B., You Z., Bryant P.J.;						
RT	"Human postsynaptic density-95 (PSD95); location of the gene (DLG4)						
RT	and possible function in nonneural as well as in neural tissues.";						
RL	Genomics 44:71-82(1997).						
RN	[2]						
RP	REVISIONS.						
RC	TISSUE=Mammary gland;						
RC	Stathakis D.G., Hoover K.H., You Z., Bryant P.J.;						
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.						
RN	[3]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=20047407; Pubmed=10582582;						
RA	Stathakis D.G., Udar N., Sandgren O., Andreasson S., Bryant P.J.,						
RA	Small K., Forsman-Semb K.;						
RT	"Genomic organization of human DLG4, the gene encoding postsynaptic						
RL	density 95.";						
RL	J. Neurochem. 73:2250-2265(1999)						

Db 217 IleAlaGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 236
QY 445 CFCATCAAGGGGTCAGCAGACAGCTCGG---CTCCAGGTAGGGAGGAGATCGTC 501
Db 237 IleIleGluGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 256
QY 502 CGGATCAATGATATTCATCTCCCTCTACCCATGAGGAGGTCAATCAACCTATTGGA 561
Db 257 AlaValAsnSerValGlyLeuGluAspValMetHisGluAspAlaValAlaLeuLys 276
QY 562 ---ACCAAGAAACGTGTCCATCAAGTG-----AGACATCGGCTGTATCCCC 609
Db 277 AsnThrTyrAspValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 296
QY 610 GTGAAAGCTCTCCGTAGCAGCCCTCACTGG---CAGTATGTGATCAGTTGTGTCG 666
Db 297 SerTyrAlaProProAspIleThrThrSerTyrSerGlnHisLeuAspAsnGluIleSer 316
QY 667 GAATCTGGGGGCTCGGAGCCAGCTGGGC----- 696
Db 317 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 332
QY 697 ---TCCCTCGAATCGGAA-----AACAG 720
Db 333 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluAspIleProArg 352
QY 721 GAGAAGAGGTCTTCATCAGCTGCTCCGAGGCTTGGGTGACAGCATTTCCAGC 780
Db 353 GluProArgArgIleValIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 372
QY 781 GGCCCCATCCAGAGCTGGCATCTTTATCAGCCATGTG-----AAACCTGGCTCC 831
Db 373 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 391
QY 832 CTCTCTGCTGAGGTGGATGAGATAGGGAGCCAGATTGTCGAGTCAATGGCTGCAG 891
Db 392 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 409
QY 892 TTCTCTCACTGGATCACAAGAGGCTGTAATGTCTGAAATAAGCCGACCTGACC 951
Db 410 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 428
QY 952 ATCTCCTATGTAGTCACTGCGGAGCTGTTTCAATGACAGACCGGAGCGGTGGCA 1011
Db 429 ValThrIleIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 443
QY 1012 GAGCGCGGCGCTGCGAGCGGAGGAGCTTCTCATCAGAGAGCGGTGGCGATG 1071
Db 444 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 461
QY 1072 GAGTCCACAAGATCCTCCAGGAGCAGCAGGAG-----ATGGAG 1110
Db 462 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 481
QY 1111 CGCAAGAGGAGAAAGAAATGCCAGAGGCGGAGAGGAGAAATGAGAGATACCGGAAG 1170
Db 482 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheArgPheGlyAsp 501
QY 1171 GAGATGGAACAGATTCTAGAGGAGGAGGAGAACTTT----- 1206
Db 502 ValLeuHisValIleAspAlaSerAspGluGluTyrTrpGlnAlaArgValHisSer 521
QY 1207 -----AAGAAGCAA 1215
Db 522 AspSerGluThrAspAspIleGlyPheIleProSerLysArgValGluArgArgGlu 541
QY 1216 TGGGAA-----GAAGACTGGGGCTCA-----AAGGAA 1242
Db 542 TrpSerArgLeuLysAlaLysAspTrpGlySerSerSerGlySerGlnGlyArgGluAsp 561
QY 1243 CAGCTACTCTTGCTAAACCATCACT---GCTGAGGTACACCCAGTACCCCTCCAG 1299
Db 562 SerValLeuSerTyrGluThrValThrGlnMetGluValHisTyrAlaArg----- 578

QY 1300 CCAAAGTGATTTCGGAATATGAGGAAGGCTTTGACCCCTACTCTATGTGTCAACCCAGA 1359
Db 579 -----ProIleIleIleLeuLeuGlyProThr 586
QY 1360 GCAGATCATGGGGAAGGATGTCCGGCTCCACGATCAAGAGGAGGATCCTTAGACCT 1419
Db 587 LysAspArgAla-----AsnAspAspLeuLeuSerGlu 597
QY 1420 GGCCCTGGAAGCGGTGTGGACTCCCCATTTGGGAAGGTGGTCTTCTGCTGTGTATGA 1479
Db 598 PheProAspLys----- 601
QY 1480 GCGGGAGCTCTGAGCGGATGTGTCATTTGTAAAGG---GGACGAGATCATGCGCAAT 1536
Db 602 PheGlySerCysValPro-----HisThrThrArgProLysArgGluTyr-GluIle 618
QY 1537 CAACGGCAAGATTGTGACAGACTACACCTGGCTGAGGCTGACGCTGCCCTGCAAGGC 1596
Db 618 eAspGlyArg-----AspTyrHisPheValSerSerArgGluLysMetGluLys-- 634
QY 1597 CTGGAATCAGCGCGGAGTGGATCGACCTTGTGTGGCTCTGCCCCCAAGAGGATA 1656
Db 634 ----- 634
QY 1657 TGACATGAGTGAACCTTTCTGCTGAAGTCCAAAGGGGAAACCAATTCACGGTTAGG 1716
Db 635 -AspIleGlnAlaHisLysPheIleGluAlaGlyGlnTyrAsnSer---HisLeuTyrGlu 653
QY 1717 AAACAGT-----GAGCTCGGCGCCACCTCGGTGAACACAAA 1752
Db 653 yThrSerValGlnSerValArgGluValAlaGluGlnGlyLysHisCysIleLeuAspVal 673
QY 1753 GCCTCGGACCGCTTGAGAGAGCCACATGACACACACAGATGGCATCTCTGGACCT 1812
Db 673 1SerAlaAsnAlaValArgArgLeuGlnAlaAlaHisLeu-----HisProIleAlaIle 691
QY 1813 GAATCTATCACCAGGAATCTCAAACTCCCTTTGGCCCTGAAC----- 1855
Db 691 ePheIleArgProArgSerLeuGluAsnValLeuGluIleAsnLysArgIleThrGluGlu 711
QY 1856 -CAGGGCCAGATAGGAACACGCTCGGCGCACTTTTTCAG- 1895
Db 711 uGln-AlaArgLysAlaPheAspArgAlaThrLysLeuGluGlnGluPheThrGluCysP 731
QY 1896 -----GCCAATGTGGAGGAGGAGGAGCAGCAGCGCTTTGGGAGAGATCAAGG---A 1947
Db 731 hSerAlaIleValGluGlyAspSerPheGluGluIleTyrHisLysValLysArgValI 751
QY 1948 TCCAGACTCTCATTT---CCTTCTCTCTGG 1973
Db 751 leGluAspLeuSerGlyProTyrIleTyr 760
RESULT 9
DLG4_RAT STANDARD; PRT; 724 AA.
ID DLG4_RAT
AC P31016; p97631;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)
DE (Synapse-associated protein 90) (Discs, large homolog 4).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=93040233; PubMed=1419001;
RA Cho K.-O., Hunt C.A., Kennedy M.B.;
RT "The rat brain postsynaptic density fraction contains a homolog of

[illegible]

Db 559 AspLeuArgGluGln-----MetMetAsnSerSerIleSerSerGlySerLeu 576
 Qy 1087 CTCCAGGAGCAGCAG 1101
 Db 577 ArgThrSerGlnLys 581

RESULT 12

DLG4_MOUSE STANDARD; PRT; 724 AA.
 ID DLG4_MOUSE
 AC Q62108;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)
 DE (Synapse-associated protein 90) (Discs, large homolog 4).
 GN DLG4 OR DLG4 OR PSD95.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DRA/2; TISSUE=Brain;
 RA Kohmura N., Yagi T.;

RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
 CC -!- SUBUNIT: Interacts with DLGAP1/GKAP (By similarity). Is part of a
 CC complex with DLGAP1/GKAP, SHANK1, SHANK2 or SHANK3 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Concentrated at synaptic
 CC junctions primarily on the presynaptic side. Also found in
 CC postsynaptic density of neuronal cells (By similarity).
 CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; D50621; BAA09297.1; --
 DR HSP; P31016; 1BE9.
 DR MGD; MGI:1277959; Dlg4.
 DR InterPro; IPR000619; Guanylate_kin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00018; SH3; 1.
 DR Pfam; PF00595; PDZ; 3.
 DR Pfam; PF00625; Guanylate_kin; 1.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00072; GuK; 1.
 DR SMART; SM00328; PDZ; 3.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS50106; PDZ; 3.
 DR PROSITE; PS50002; SH3; 1.
 DR SH3 domain; Repeat.
 FT DOMAIN 65 151 PDZ 1.
 FT DOMAIN 160 246 PDZ 2.
 FT DOMAIN 313 393 PDZ 3.
 FT DOMAIN 428 498 SH3.
 FT DOMAIN 534 724 GUANYLATE KINASE.
 SQ SEQUENCE 724 AA; 80472 MW; 7EFFC99E1FF90BA CRC64;

Alignment Scores:

Pred. No.: 5,46e-06 Length: 724
 Score: 229,50 Matches: 146

Percent Similarity: 37.54% Conservative: 101
 Best Local Similarity: 22.19% Mismatches: 204
 Query Match: 5.75% Indels: 207
 DB: 1 Gaps: 35

US-09-502-945-5 (1-2162) x DLG4_MOUSE (1-724)

Qy 340 TCCAGGAAGCTGAAGGAGTGGCTCTGGACCGCTGACCCGAGGCGCTCGGCTGAGT 399
 Db 155 AlaGluLysIleIleGluIleLysLeuIleLys---GlyProLysGlyLeuGlyPheSer 173
 Qy 400 GTGGCTGTGGTGGC-----CTGGAGTTGGCTGGCTCTTCACTCCAC 444
 Db 174 IleAlaGlyGlyValGlyAsnGlnHisIleProGlyAspAsnSerIleTyrValThrLys 193
 Qy 445 CTCATCAAGCGCGTCAGGACGACAGCGTCGGG---CTCCAGGTAGGGAGCAGATCGTC 501
 Db 194 IleIleGluGlyGlyAlaAlaHisLysAspGlyArgLeuGlnIleGlyAspLysIleLeu 213
 Qy 502 CGGATCATGATATTCCATCTCTCTGTACCCATGAGGAGGTCATCAACCTCATTCGA 561
 Db 214 AlaValAsnSerValGlyLeuGluAspValMethHisGluAspAlaValAlaLeuLys 233
 Qy 562 ---ACCAAGAAACTGTGTCCATCAAAAGTG-----AGACACATCGGCTCATCCCC 609
 Db 234 AsnThrTyrAspValValTyrLeuLysValAlaLysProSerAsnAlaTyrLeuSerAsp 253
 Qy 610 GTGAAAAGCTCTCTGTAGAGCCCTCCTCTGG---CAGTATGTGGATCAGTTGTGTCTCG 666
 Db 254 SerTyrAlaProAspIleThrThrSerTyrSerGlnHisLeuAspAsnGluIleSer 273
 Qy 667 GAATCTGGGCGGTGCGAGGCGCTGGGC----- 696
 Db 274 HisSer-----SerTyrLeuGlyThrAspTyrProThrAlaMetThrProThr 289
 Qy 697 -----TCCCTCGAAATCGGAA-----AACAG 720
 Db 290 SerProArgArgTyrSerProValAlaLysAspLeuLeuGlyGluGluAspIleProArg 309
 Qy 721 GAGAAGAGGTTCATCAGCGCTGGTAGCTCCGAGGCGCTGGCTGCAGCATTTCCAGC 780
 Db 310 GluProArgIleIleHisArgGlySerThrGlyLeuGlyPheAsnIleValGly 329
 Qy 781 GGCCCCATCAGAGCGCTGCATCTTTATCAGCCATGT-----AACTGGGTCTCC 831
 Db 330 GlyGlu---AspGlyGluGlyIlePheIleSerPheIleLeuAlaGlyGlyProAlaAsp 348
 Qy 832 CTGTCTCTGTAGGTGGAGTGGAGATAGGGGACAGATTCGAAAGTCAATGGCTGCAC 891
 Db 349 LeuSerGlyGlu-----LeuArgLysGlyAspGlnIleLeuSerValAsnGlyValAsp 366
 Qy 892 TTCTCTAACCTGGATCACAGGAGGCTGTAAATGTGCTGAAAATAGCCGACGCTGACC 951
 Db 367 LeuArgAsnAlaSerHisGluGlnAlaAlaIleAlaLeuLysAsnAla---GlyGlnThr 385
 Qy 952 ATCTCCATTGTAGCTGCAGCTGGCCGGGAGCTGTTTCATGACAGACCGGGAGCGCTGGCA 1011
 Db 386 ValThrIleIleAlaGln-----TyrLysProGluGluTyrSerArgPhe 400
 Qy 1012 GAGCGCGCGAGCGTGTAGCTGCAGCGGACGAGGTCTTCATCGAAGAGCGGCTG----- 1065
 Db 401 GluAlaLysIleHisAspLeuArgGluGln-----LeuMetAsnSerSerLeuGlySer 418
 Qy 1066 -----CGCATGGAGTCCAAC----- 1080
 Db 419 GlyThrAlaSerLeuArgSerAsnProLysArgGlyPheTyrIleArgAlaLeuPheAsp 438
 Qy 1080 ----- 1080
 Db 439 TyrAspLysThrLysAspCysGlyPheLeuSerGlnAlaLeuSerPheHisPheGlyAsp 458
 Qy 1081 -----AAGATCTCTCCAGGAGCAGCAGGACATGGAGCGGCAAGAGGAGAAATGTC 1134
 Db 1134


```
Db 459 ValLeuHisValIleAspAlaSerAspGluGluTrpTrpGlnAlaArgValHisSer 478
QY 1135 CAGAAGGAGCAGCAAAATGAGAGATACCGGAAGGAGATGGAACACAGATGTAGAGGAG 1194
Db 479 AspSerGluThrAsp-----AspIleGlyPheIleProSerLys 491
QY 1195 GAAGAGAAGTTTAAGAACAATGGAA-----GAAGACTGGGGCTCA----- 1236
Db 492 ArgArgValGluArgArgGluTrpSerArgLeuLysAlaLysAspTrpGlySerSer 511
QY 1237 -----AAGAACAGACTACTCTGTCCTCAAAACCATCACT---GCTGAGGTA 1278
Db 512 GlySerGlnGlyArgGluAspSerValLeuSerTyGluThrValThrGlnMetGluVal 531
QY 1279 CACCAGTACCCCTTCGCAAGCAAAAGTATTCCGGAAATATGAGGAAGGCTTTGACCC 1338
Db 532 HisTyAlaArg-----Pro 536
QY 1339 CTACTCTATGTTTACCCAGACAGATCATGCGGAAGGATGTCGGCTCCTACGCATCAA 1398
Db 537 IleIleIleLeuGlyProThrLysAspArgAla----- 547
QY 1399 GAAGGAGGATCCTTAGACCTGGCCTGGAAGCGGTGTGGACTCCCCCATTTGGAAGGT 1458
Db 548 AsnAspAspLeuLeuSerGluPheProAspLys----- 558
QY 1459 GTCGCTTCTGCTGTATGATGCGGGGAGCTGCTGAGCGGCATGTGGCATTTGAAAGG 1518
Db 559 -----PheGlySerCysValPro-----HisThrThrArg 568
QY 1519 ---GGACGAGATCATGGCAATCAACGCGCAAGATTGTGACAGACTACACCTGGCTGAGGC 1575
Db 569 ProLysArgGluTyrr-GluIleaspGlyArg-----AspTyrrHisPheValSerSe 585
QY 1576 TGACGCTGCCCTGCAGAAAGCCCTGGAAATCAGGCGGGGACTGGATGACCTGTGTGGTGC 1635
Db 585 rArgGluLysMetGluLys----- 591
QY 1636 CGTCTCCCCCAAGGAGTATGACGATGACGTGACCTCTTCTGTAAGTCCAAAAGGGG 1695
Db 592 -----AspIleGlnAlaHisLysPheIleGluAlaGlyGlnTy 604
QY 1696 AACCAAAATTCACGCGTTAGGAACAGT-----GAGCTCCG. 1731
Db 604 rAsnSer---HisLeuTyrrGlyThrSerValGlnSerValArgGluValAlaGluGlnG 623
QY 1732 GCCCCACCTCGTGAACACAAAGCCTCGGACCCTTGAGAGAGGCCACATGACACACAC 1791
Db 623 yLysHisCysIleLeuAspValSerAlaAsnAlaValArgLeuGlnAlaAlaHisLe 643
QY 1792 CAGATGCCATCCTTGGGACTGAATCTATCACCAGGAATCTCAAACTCCCTTTGGCCCT 1851
Db 643 u-----HisProIleAlaIlePheIleArgProArgSerLeuGluAsnValLeuGluI 661
QY 1852 GAAC-----CAGGCCAGATAAGAAACAGCTCGGGCCACTTTTTCGA 1893
Db 661 eAsnLysArgIleThrGluGluGln-AlaArgLysAlaPheAspArgAlaThrLysLeuG 681
QY 1894 AG-----GCCAATGTGAGGAAGGAGGAGCCAGCCGCTTT 1929
Db 681 luGlnGluPheThrGluCysPheSerAlaIleValGluGlyAspSerPheGluGluIle 701
QY 1930 GGGAGAGAGACTCAAGG---ATCCAGACTCTCATT---CCTTTCCCTCTGG 1973
Db 701 yrHisLysValLysArgValIleGluAspLeuSerGlyProTyrrIleTrp 717
```

RESULT 13

```
DLGL_RAT STANDARD; PRT; 911 AA.
ID DLGL_RAT
AC O62656;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

Alignment Scores:

```
Pred. No.: 9,04e-06 Length: 911
Score: 226.00 Matches: 120
Percent Similarity: 37.32% Conservative: 83
Best Local Similarity: 22.06% Mismatches: 180
Query Match: 5.67% Indels: 161
DB: 1 Gaps: 24
```

```
DE Presynaptic protein SAP97 (Synapse-associated protein 97) (Discs,
GN large homolog 1).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95198112; PubMed=7891172;
RA Mueller B.M., Kistner O., Veh R.W., Cases-Langhoff C., Becker B.,
RA Gundelfinger E.D., Garner C.C.;
RT "Molecular characterization and spatial distribution of SAP97, a
RT novel presynaptic protein homologous to SAP90 and the Drosophila
RT discs-large tumor suppressor protein.";
RL J. Neurosci. 15:2354-2366(1995)
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTogenesis. MAY PLAY A ROLE IN
CC CELL ADHESION.
CC -!- SUBCELLULAR LOCATION: IN CULTURED T84 CELLS, IT IS RESTRICTED TO
CC THE CYTOPLASMIC SURFACE OF THE PLASMA MEMBRANES BETWEEN ADJACENT
CC CELLS, BUT NOT AT THE EDGES OF CELLS LACKING CELL-CELL CONTACT.
CC -!- TISSUE SPECIFICITY: PRESYNAPTIC NERVE TERMINI OF EXCITATORY
CC SYNAPSES. IN OTHER BRAIN REGIONS, SAP97 IS FOUND IN AND ALONG
CC BUNDLES OF UNMYELINATED AXONS. SAP97 IS NOT RESTRICTED TO THE CNS,
CC BUT IS ALSO PRESENT AT THE BASAL LATERAL MEMBRANE BETWEEN A
CC VARIETY OF EPITHELIAL CELLS.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14950; AAA79976.1; -.
CC DR HSSP; Q12959; IPDR.
CC DR InterPro; IPR000619; Guanylate_kin.
CC DR InterPro; IPR001478; PDZ.
CC DR InterPro; IPR001452; SH3.
CC DR Pfam; PF00018; SH3; 1.
CC DR Pfam; PF00595; PDZ; 3.
CC DR Pfam; PF00625; Guanylate_kin; 1.
CC DR ProDom; PD000066; SH3; 1.
CC DR SMART; SM00072; GuK; 1.
CC DR SMART; SM00228; PDZ; 3.
CC DR SMART; SM00326; SH3; 1.
CC DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
CC DR PROSITE; PS00856; GUANYLATE_KINASE_2; 1.
CC DR PROSITE; PS0106; PDZ; 3.
CC DR PROSITE; PS00002; SH3; 1.
CC DR SH3 domain; Repeat.
CC FT DOMAIN 224 310 PDZ 1.
CC FT DOMAIN 318 404 PDZ 2.
CC FT DOMAIN 465 545 PDZ 3.
CC FT DOMAIN 580 650 SH3.
CC FT DOMAIN 721 911 GUANYLATE KINASE.
CC FT DOMAIN 527 530 POLY-ALA.
CC SQ SEQUENCE 911 AA; 100570 MW; 18CEBD31DD0CAF8B CRC64;
```


[illegible]

Db	645	ArgValGluLysLysGluArgAlaArgLeu-----LysThrValLysPheAsnSerLys	662
QY	1162	TACCAGGAGGAGTGAACAGATTGTAGAGGAGGAGAAAGACACT-----TT	1206
Db	663	ThrArgLysAspLysGlyGluIleProAspAsp-MetGlySerLysGlyLeuLysHisVa	682
QY	1207	AAGAGCAATGGGAAGAACTGGGCTCAAGAGAACACTACTCTTGCTCAAAACCATC	1266
Db	682	IthrSerAsnAlaSerAspSerGluSerSerTyrHisGluTyrGlyCysSerLysGlyG	702
QY	1267	ACTGCTGGGTACACC---CAGTACCCCTTCGACGCCAA-----	1303
Db	702	YgluGluGluTyrValLeuSerTyrGluProValAsnGlnGluValAsnTyrThr	722
QY	1304	-----AGTGATT	1311
Db	722	gProValIleIleLeuGlyProMetLysAspArgValAsnAspAspLeuLeuSerGluPh	742
QY	1312	C-----CGAAATATGAGGA	1326
Db	742	eProAspLysPheGlySerCysValProHisThrThrArgProLysArgAspTyrGluVa	762
QY	1327	AGGCTTTGACCCCTACTCTATGTCACCCAGACAGATCATGGGAAGAGTATGTCGGCT	1386
Db	762	lasGlyArgAspTyrHisPheValThrSerArgGluGlnMetGluLysAspIleGlnG	782
QY	1387	CCTACGCATCAAGAAGGAGGATCTTACACCTGGCCCTGGAAGC---GGTGGGCACT	1443
Db	782	uHisLysPheIleGluAlaGlyIleTyrAsnAsnHisLeuTyrGlyThrSerValGlnSe	802
QY	1444	CCCATTGGGAAGTGCTGTTCTGCTGTATGAGCGGGAGCTGCTGACGCGCATGG	1503
Db	802	r-----ValArgAlaValAlaGluLysGly-----LysHisCys	813
QY	1504	TGGCATTGTGAAGGGACGAGCATGCGCATCAACGGCAGATGTGCACAGACTACAC	1563
Db	813	s-----IleLeuAspValSerGlyAsnAlaIleLysArgLeuG	826
QY	1564	CCTGGCTGAG	1573
Db	826	nIleAlaGln	829
RESULT	14		
ZOL_HUMAN			
ID	ZOL_HUMAN	STANDARD;	PRT; 1736 AA.
AC	Q07157;		
DT	01-OCT-1994	(Rel. 30, Created)	
DT	01-OCT-1994	(Rel. 30, Last sequence update)	
DT	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Tight junction protein ZO-1 (zonula occludens 1 protein) (zona occludens 1 protein) (tight junction protein 1).		
DE	GN1 OR ZOL.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RA	MEDLINE=93361541; PubMed=8395056;		
RX	Willott E., Balda M.S., Fanning A.S., Jameson B., van Itallie C.,		
RA	Anderson J.M.;		
RT	"The tight junction protein ZO-1 is homologous to the Drosophila		
RT	discs-large tumor suppressor protein of septate junctions.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 90:7834-7838 (1993).		
CC	-1- FUNCTION: THE N-TERMINAL MAY BE INVOLVED IN TRANSDUCING A SIGNAL		
CC	REQUIRED FOR TIGHT JUNCTION ASSEMBLY, WHILE THE C-TERMINAL MAY		
CC	HAVE SPECIFIC PROPERTIES OF TIGHT JUNCTIONS. THE ALPHA DOMAIN		
CC	MIGHT BE INVOLVED IN STABILIZING JUNCTIONS.		
CC	-1- SUBUNIT: INTERACTS WITH OCCLUDIN, CLAUDINS AND ZO-3.		
CC	-1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE. CYTOPLASMIC SIDE.		
CC	MOVEMENT OF ZO-1 FROM THE CYTOPLASM TO MEMBRANE IS AN EARLY EVENT		
CC	OCCURRING CONCURRENTLY WITH CELL-CELL CONTACT		

QY 1373 -----AAGATGTCGGCTCCTACGATCAAGAAGGAGG 1406
Db 399 InLeuMetLysMetGlyPheLeuArgProSerMetLysLeuValLysPheArgLysGlyA 419
QY 1407 GATCCTTAGACCTGGCCCTGGAGCGGTGTGGACTCCCCATTTGGGAAGGTGGCTGTTT 1466
Db 419 spSerValGlyLeuArgLeuAlaGlyGlyAsnAsp-----ValGly---IlePheValA 436
QY 1467 CTGCTGTATGAGCGGGAGCTGCTGAGCGGCATGGTGGCTATTTGAAGGGGACGAGA 1526
Db 436 laGlyValLeuGluAspSerProAlaAlaLysGlu---GlyLeuGluGluGlyAspGlnI 455
QY 1527 TCATGCCAATCAAGCGCAAGATGTGACAGACTACACCTGGCTGAGCGTGCAGCTGCC 1586
Db 455 leLeuArgValAsnAsnValAspPheThrAsnIleileArgGluGluAlaValLeuPheL 475
QY 1587 TCGAGAAGCCCTGGAATCATCGCGGGAGCTGCATCGACCTTGTGGTTCGCTGCCGCC 1646
Db 475 eu-----LeuAspLeu-----p 479
QY 1647 CAAGGAGTATGACGATGAGCTGACCTTCTGCTGAAGTCCAAAGGGGAAACCAATTC 1706
Db 479 roLys-----GlyGluGluValThrIleLeuAlaGlnLysLysLys----- 492
QY 1707 AGCGCTTAGGAACAGTGCAGCTCGCGCCGCCCTGCTGTAACACACAAAGCTCGGACGCC 1766
Db 493 -----AspValTyArgArgIleValGluSerAspValGlyAspSerp 507
QY 1767 TTGAG---AGAGGCCACATGACACACACAGATGCATCCTTGGGACCTGAATCTATCAC 1823
Db 507 heTyriLeargThrHisPheGluTyriGluLysGluSerProTyrgly----- 522
QY 1824 CCAGGAATCTCAACCTCCTTGGCCCTGAACACGCGGCAGATAAGGAACAGCTCGGGCC 1883
Db 523 -----LeuSerPheAsnLysGlyGluValPheArgAlaValAsp 536
QY 1884 ACTTTTGAAGGCCAATGTGGAGGAAGGGAGCAGCGCTTTGGGAGAAGATCTCA 1943
Db 536 hrLeuTyriAsnGlyLysLeu---GlySerTrpLeuAlaIleArgIleGlyLysAsnHisL 555
QY 1944 AGGATCCAGACTCTCATCTTCTCT-----CTGGCCCAAGTGAA 1982
Db 555 ysgluValGluArgGlyIleleProAsnLysAsnArgAlaGluGlnLeuAlaSerValG 575
QY 1983 TTGCTCTCTCCCA-----GCTTGGGGGAC 2008
Db 575 InTyriThrLeuProLysThrAlaGlyGlyAsp 585

RESULT 15
ID DLG2_RAT STANDARD; PRT; 852 AA.
AC Q63622; Q62939; P70548;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Channel associated protein of synapse-110 (Chapsyn-110) (Synaptic
DE density protein PSD-93) (Discs, large homolog 2).
GN DLG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96310881; PubMed=8755482;
RA Kim E., Cho K.O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of
RL Chapsyn-110, a member of the PSD-95 family of proteins.";
RN Neuron 17:103-113(1996).
RP SEQUENCE FROM N.A.
RX MEDLINE=96193770; PubMed=8625413;
RA Brennan J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E.,

Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F.,
Froehner S.C., Bredt D.S.;
"Interaction of nitric oxide synthase with the postsynaptic density
protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
Cell 84:757-767(1996).
[3]
RP SEQUENCE FROM N.A.
RA Irie M., Hata Y., Takai Y.;
Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS AS WELL AS POTASSIUM CHANNELS.
CC SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC SIMILARITY: CONTAINS 1 GUANYLATE KINASE-LIKE DOMAIN.
CC SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; U49049; AAB53243.1; -;
EMBL; U50717; AAC52643.1; -;
EMBL; U53368; AAB48562.1; -;
HSSP; Q12959; IPDR.
InterPro; IPR000619; Guanylate_kin.
InterPro; IPR001478; PDZ.
InterPro; IPR001452; SH3.
Pfam; PF00018; SH3; 1.
Pfam; PF00595; PDZ; 3.
Pfam; PF00625; Guanylate_kin; 1.
ProDom; PD000066; SH3; 1.
SMART; SM00072; GuK; 1.
SMART; SM00228; PDZ; 3.
SMART; SM00326; SH3; 1.
PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
PROSITE; PS0106; PDZ; 3.
PROSITE; PS00002; SH3; 1.
SH3 domain; Repeat.
FT DOMAIN 98 184 PDZ 1.
FT DOMAIN 193 279 PDZ 2.
FT DOMAIN 421 501 PDZ 3.
FT DOMAIN 536 606 SH3.
FT DOMAIN 662 852 GUANYLATE KINASE.
FT CONFLICT 181 182 VR -> IL (IN REF. 2).
FT CONFLICT 228 228 I -> M (IN REF. 2).
FT CONFLICT 326 326 R -> K (IN REF. 2).
FT CONFLICT 339 339 D -> E (IN REF. 3).
FT CONFLICT 450 454 MISSING (IN REF. 2).
FT CONFLICT 464 465 GD -> RK (IN REF. 2).
FT CONFLICT 474 474 D -> H (IN REF. 2).
FT CONFLICT 476 476 R -> P (IN REF. 2).
FT CONFLICT 478 478 A -> D (IN REF. 2).
FT CONFLICT 484 486 AAA -> LP (IN REF. 2).
FT CONFLICT 506 506 A -> S (IN REF. 2).
FT CONFLICT 569 569 H -> N (IN REF. 2).
FT CONFLICT 586 586 L -> Q (IN REF. 2).
FT CONFLICT 626 641 MISSING (IN REF. 2).
FT CONFLICT 639 639 K -> A (IN REF. 3).
FT CONFLICT 726 726 F -> L (IN REF. 1).
FT CONFLICT 733 733 N -> Y (IN REF. 2).
FT CONFLICT 749 749 E -> V (IN REF. 1).
FT CONFLICT 756 756 L -> H (IN REF. 2).
FT CONFLICT 791 792 KR -> NG (IN REF. 2).
FT CONFLICT 794 794 T -> M (IN REF. 2).
SQ SEQUENCE 852 AA; 94934 MW; F8D414A8B9CF5B09 CRC64;

Alignment Scores: 3.58e-05 Length: 852
Pred. No.:

Score: 215.50 Matches: 119
Percent Similarity: 39.50% Conservative: 86
Best Local Similarity: 22.93% Mismatches: 188
Query Match: 5.40% Indels: 126
DB: 1 Gaps: 24

US-09-502-945-5 (1-2162) x DLG2_RAT (1-852)

```
QY 307 CAGGTGGGAATATGATCAGCTGACCCCGCGCTCCAGAGCTCAAGGAGGTGGCTGTG 366
Db 91 GllleGluGlyGlu-----PheGluGluIleThrLeu 101
QY 367 GACCGTCTGCACCCCAAGGCTCGGCTGAGTGTGCGTGGCTGGAG----- 417
Db 102 GluArgGlyAsn---SerGlyLeuGlyPheSerIleAlaGlyGlyThrAspAsnProHis 120
QY 418 -----TTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGGGCGGTGACGACAGC 471
Db 121 IleGlyAspProGlyIlePheIleThrLysIlePheGlyGlyAlaAlaGlu 140
QY 472 GTCGGG---CTCCAGGTAGGGGAGATCGTCGGGATCAATGGATATTCATCTCTCC 528
Db 141 AspGlyArgLeuArgValAsnAspCysIleLeuArgValAsnGluValAspValSerGlu 160
QY 529 TGTACCCATGAGGAGTCAACCTCATTCGA---ACCAAGAAACTGTGTCCATCAA 585
Db 161 ValSerHisSerLysAlaValGluAlaLeuLysGluAlaGlySerIleValArgLeuTyr 180
QY 586 GTCGACACATCGGCTGATCCCGTGAAGCTCTCTGATGAGCCCTCCTACTGGCAG 645
Db 181 ValArgArg----- 183
QY 646 TATGTGGATCAGTTTGTGTGGAATCTGGGGGTGCGGAGCGCTGGCTCCCTGGA 705
Db 183 ----- 183
QY 706 AATCGGAAACAGAGAGAGAGTCTTCATCAGCCTGGTA---GGCTCCCGAGCCCTT 762
Db 184 ArgArgProIleLeuGluThrValValGluIleLysLeuPheLysGlyProLysGlyLeu 203
QY 763 GGCTGCAGGATTCACGGCGCCCATCCAGAG---CCTGGC-----ATCTTT 807
Db 204 GlyPheSerIleAlaGlyValGlyValGlyValGlyValGlyValGlyValGlyValGly 223
QY 808 ATCAGCATGTGAACCTGCTCCCTGCTGCTGAGTGGGA---TTGGAGATAGGGGAC 864
Db 224 ValThrLysIleLeuAspGlyGlyAlaAlaGlnLysAspGlyArgLeuGlnValGlyAsp 243
QY 865 CAGATTGCTGAAGTCAATGGCGTGGACTTCTTAACCTGGATCACAGGAGGCTGTAAT 924
Db 244 ArgLeuLeuMetValAsnAsnTyrSerLeuGluGluValThrHisGluGluAlaValAla 263
QY 925 GTGCTGAATAATAGCCCGACCTGACCATCTCCATCTGCTGAGTGGCGCGGAGCTG 984
Db 264 IleLeuLysAsnThr---SerAspValValThrLysValGlyLysProThrThrIle 282
QY 985 TTCATGACAGACCGGGGCTGGCAGAGCGCGGAGCGGTGAGTGTGACGCTGGCGGAGGAG 1044
Db 283 TyrMetThrAspProTyrGlyProProAspIle-----ThrHisSerTyrSer 298
QY 1045 CTTCTCATGAGAGCGGCTGGGATGGATGAGTCCCAAGATCTCCAGGAGCAGCAGGAG 1104
Db 299 ProProMetGluAsnHisLeuLeuSerGlyAsnAsnGlyThrLeuGluTyrLysThrSer 318
QY 1105 ATG-----GAGCGGCAAGGAGAAAGAAATTCGCCAGAGGCGCAGAGGAAAT 1155
Db 319 LeuProIleSerProGlyArgTyrSerProIleProLysHisMetLeuValGluAsp 338
QY 1156 GAGAGATACCGGAGGAGATGGATCTTAGAGGAGGAGAGAGTAAAGAACAA 1215
Db 339 Asp---TyrThrArgProProGluProValTyrSerThrValAsnLys-LeuCysAsp-- 356
QY 1216 TGGGAAGAAGACTGGGGCTCAAGGAGACAGCTACTCT-----TGCCTAAACCATC 1266
```

```
Db 357 ----LysProAlaSerProArgHis---TyrSerProValGluCysAspLys---Se 372
QY 1267 ACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATTCCGGAATATGAGGA 1326
Db 372 rPheLeuLeuSerThrProTyrPro-----HisTyrHisLe 384
QY 1327 AGGTTTGGACCCCTACCTCTATGTTTACCCCCAGAG----- 1360
Db 384 uGlyLeuLeuProAspSerAspMetThrSerHisSerGlnHisSerThrAlaThrArgGl 404
QY 1361 -----CAGATCATGGGGAAGGATGTCGGCTCCCTACG 1392
Db 404 nProSerValThrLeuGlnArgAlaIleSerLeuLeuGlyGluProArgLysValValLe 424
QY 1393 CATCAAGAAGGAGGATCCTTAGACCTGGCCCTGGAAGCGGTGTGGACTCCCCCATTTGG 1452
Db 424 uHisLysGlySerThrGlyLeuGlyPheAsnIleValGlyGlyGluAspGlyGlu----- 442
QY 1453 GAAGTGTCTGTTCTGCTGTGTATGACGGGGAGCTCTGAGCGGCATGGTGGCATTTGT 1512
Db 443 -GlyIlePheValSerPheIleLeuAlaGlyGlyProAlaAspLeuSerGlyGluLeuGl 462
QY 1513 GAAAGGGGACGAGATCATGGCAATCAACGGCAAGATTGTGACAGACTACACCTGGCTGA 1572
Db 462 nArgGlyAspGlnIleLeuSerValAsnGlyIleAspLeuArgGlyAlaSerHisGluGl 482
QY 1573 GGCTGACCTGCTCCCTGTCAGAGGCTGGAATCAGGCGGGGAGTGGATCGACCTGTGTGT 1632
Db 482 nAlaAlaAlaAlaLeu-----LysGlyAlaGlyGlnThrValThrIleIl 497
QY 1633 TCCTCTGCTGCCCCCAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1692
Db 497 eAlaGlnTyrGlnProGluAspTyr-----AlaArgph 508
QY 1693 GGGAAACCAATTCACGCTTAGGAAACAGTGCAGTCCGCCCCACCTCGTGAAC 1747
Db 508 eGluAlaLysIleHis-----AspLeuArgGluGlnMetMetAsn 521
```

Search completed: March 21, 2003, 12:48:39
Job time : 51.5761 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:20:42 ; Search time 74.1434 Seconds
(without alignments)
12016.549 Million cell updates/sec

Title: US-09-502-945-5

Perfect score: 3989

Sequence: 1 cctggccggctgcgctgcg.....tatttccagcttaaaaaa 2162

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09502945/runat_14032003_101058_19113/app_query.fasta_1.10379
-DB-SPTREMBL_21 -QFMT=fastan -MATRIX=spool_0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -SUFFIX=rspt -MINMATCH=0.1 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pcpt -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09502945.ecgn_1.1.565.ernat_14032003_101058_19113 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21.*
1: sp.archaea.*
2: sp.bacteria.*
3: sp.fungi.*
4: sp.human.*
5: sp.invertebrate.*
6: sp.mammal.*
7: sp.mhc.*
8: sp.organella.*
9: sp.phage.*
10: sp.plant.*
11: sp.potent.*
12: sp.virus.*
13: sp.vertebrate.*
14: sp.unclassified.*
15: sp.virus.*
16: sp.bacteriap.*
17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2558.5	64.1	548	11	Q9ES65	Q9es65 mus musculus

ID	Q9ES65	PRELIMINARY	PRT	548 AA
AC	Q9ES65			
DT	01-MAR-2001 (TREMBlrel. 16, Created)			
DT	01-MAR-2001 (TREMBlrel. 16, Last sequence update)			
DE	01-DEC-2001 (TREMBlrel. 19, Last annotation update)			
DE	Harmonin isoform al.			
GN	2010016F01RIK OR USHIC.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090,			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20428180; PubMed=10973247;			
RA	Verpy E., Leibovici M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,			
RA	Mansour A., Blanchard S., Kobayashi I., Keats B.J., Slim R., Petit C.,			
RT	"A defect in harmonin, a PDZ domain-containing protein expressed in			
RT	the inner ear sensory hair cells, underlies Usher syndrome type 1C.";			
RL	Nat. Genet. 26:51-55(2000).			
DR	EMBL; AF228924; AAG12457.1; -			
DR	HSSP; P29476; IOAV.			
DR	MGD; MGI:1919338; 2010016F01RIK.			
DR	InterPro; IPR001478; PDZ.			

ALIGNMENTS

RESULT 1

Q9ES65
ID Q9ES65 PRELIMINARY PRT 548 AA

AC Q9ES65
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Harmonin isoform al.
GN 2010016F01RIK OR USHIC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090,
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20428180; PubMed=10973247;
RA Verpy E., Leibovici M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,
RA Mansour A., Blanchard S., Kobayashi I., Keats B.J., Slim R., Petit C.,
RT "A defect in harmonin, a PDZ domain-containing protein expressed in
the inner ear sensory hair cells, underlies Usher syndrome type 1C.";
RL Nat. Genet. 26:51-55(2000).
DR EMBL; AF228924; AAG12457.1; -
DR HSSP; P29476; IOAV.
DR MGD; MGI:1919338; 2010016F01RIK.
DR InterPro; IPR001478; PDZ.

DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS50106; PDZ; 3
SQ SEQUENCE 548 AA; 6171

Alignment Scores:

Pred. No.:	3.4e-177	Length:	548
Argument Scores:		Matches:	503
Score:	2558.50	Conservative:	12
Percent Similarity:	93.98%	Mismatches:	11
Best Local Similarity:	91.79%	Indels:	22
Query Match:	64.14%	Gaps:	1
DB:	11		

US-09-502-945-5 (1-2162) x 09ES65 (1-548)

Qy	97	ATGACCGCAAAGTGGCCCGAGAAATTC	CGGCATAAAGTGGATTTCTGTGATGAAATGAT	156
Db	1	MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp		20
Qy	157	GCAGAGAAGCAGCATCTCTATGATGTGCTCGCAATCTACCAACAGACCATGACGCTGGCC	216	
Db	21	AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla	40	
Qy	217	GTGCTCTGGGAGACCTGAAGCTGGTCATCAATGAACCCAGCCGCTGCCTCTGTTTGAAT	276	
Db	41	ValLeuValGlyAspLeuLysLeuValIleAsnGluProAsnArgLeuProLeuPheAsp	60	
Qy	277	GCCATTGGCGCGCTGATCCCACTAAGCACCAGCGTGAATATCATCAGCTGACCCCGG	336	
Db	61	AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg	80	
Qy	337	CGCTCCAGGAAGCTGAAGGAGTGCCTCGACCGTCTGCACCCCAAGGCGCTCGGCCGTG	396	
Db	81	ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu	100	
Qy	397	AGTGTGCGTGTGGCCCTGGAGTTTGCGTGTGGCTCTTCATCTCCACCTCATCAAAGC	456	
Db	101	SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly	120	
Qy	457	GGTCAGCGAGACAGCTGCGGGCTCCAGGTAGGGAGCAGAGATCTCGGGATCAATGATAT	516	
Db	121	GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr	140	
Qy	517	TCCATCTCTCTCTACCCATGAGGAGGTTCATCAACCTCATTCGAACCAAGAAACTGTG	576	
Db	141	SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal	160	
Qy	577	TCCATCAAGTGGAGACATCGGCCTGATCCCGTGAAGAGCTCTCCTGTATGAGCCCTTC	636	
Db	161	SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProGluGluSerLeu	180	
Qy	637	ACTTGGCAGTATGTGGATCAGTTTCTCGGAATCTCGGGCGCTGGAGGCAGCCCTGGGC	696	
Db	181	LysTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlyGlyLeuGly	200	
Qy	697	TCCCTCGAAATCGGAAAAACAAGAGAGCTTCATCAGCTGTGAGGCTCCCGA	756	
Db	201	SerProGlyAsnArgThrThrLysGluLysLysValPheIleSerLeuValGlySerArg	220	
Qy	757	GGCCTTGGCTGCAGCATTTCCAGCGGGCCCCATCCAGAAGCCTTGGCATCTTTATCAGCCAT	816	
Db	221	GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheValSerHis	240	
Qy	817	GTGAACCTGGCTCCCTCTCTCTGCTGAGGTGGGATGGGAGCAGATGTGCGAA	876	
Db	241	ValLysProGlySerLeuSerAlaGluValGlyLeuGluThrGlyAspGlnIleValGlu	260	
Qy	877	GTCAATGGCTCGCATCTCTAACTGGATCACAAGGAGGCTGTAAATGTGCTGAAAAAT	936	
Db	261	ValAsnGlyIleAspPheThrAsnLeuAspHisLysGluAlaValAsnValLeuLysSer	280	
Qy	937	AGCCGCGACCTGACCATCTCCATTGTAGCTGCATCGCCGGGAGCTGTTCATGCACGAC	996	

D	b	281	SerArgSerLeuThrIleSerIleValAlaGlyAlaGlyArgGluLeuPheMetThrAsp	300
Q	y	997	CGGGAGCGCTGGCAGAGCGCGGCAGCGTGCAGTCGACGGCAGGAGCTTCATGTCAg	1056
D	b	301	ArgGluArgLeuGluAlaAargGlnArgGluLeuGlnArgGlnGluLeuLeuMetGln	320
Q	y	1057	AAGCGCTGGCATGGAGTCCAAACAAGATCTCCAGGAGCAGCAGGAGATGGAGCGCAA	1116
D	b	321	LysArgLeuAlaMetGlnSerAsnLysIleLeuGlnGlnGlnGlnGluMetGluArgGln	340
Q	y	1117	AGGAGAAAAGAATAATGCCAGGAAGCAGCAGAGGAAAAATCAGAGATACCAGGAAGGAGATG	1176
D	b	341	ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet	360
Q	y	1177	GAAcAGATTGTTAGAGAGGAAGAAGTTTAAAGAACATGGGGAAGAAGACTGGGGCTCA	1236
D	b	361	GluGlnIleSerGluGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer	380
Q	y	1237	AAGGAACAGCTACTCTTGCGCTAAACACCATCACTGCTGAGGTFACACCAGTACCCCTTCGC	1296
D	b	381	LysGluGlnLeuIleLeuProLysThrIleThrAlaGluValHisProValProLeuArg	400
Q	y	1297	AAAGCCA-----	1302
D	b	401	LysProLysTyrAspGlnGlyValGluProAlaAspHisLeuAspGlySerThrGluGlu	420
Q	y	1303	-----AAGTGATTCCGGAAATATAGGAAGGCTTTCACCCCCTACTCTATGTTCAc	1353
D	b	421	GlnArgGlnGln- AspPheArgLysTyrGluGluGlyPheAspProTyrSerMetPheSe	440
Q	y	1354	CCcAGACAGATCATCGGGAAGGAGTTCGGGCTCCTCAGCATCAAGAAAGAGGAGGATCCTT	1413
D	b	440	rProGluGlnIleAlaGlyLysAspValArgLeuLeuArglleLysLysGluGlySerLe	460
Q	y	1414	AGAcCTGGCCCTGGAAGGGGGTGTGGACTCCCCATTGGGAAGGTGGTCTTCGTGT	1473
D	b	460	uasPLeuAlaLeuGluGlyGlyValAspSerProValGlyLysValValSerAlaVa	480
Q	y	1474	GTATAGCGGGGAGCTGTGAGCGGCATGGTGGCAATTGTGAAGGGGACGAGATCATGGC	1533
D	b	480	lTyrGluGlyGlyAlaAlaGluArgHisGlyGlyValLysGlyAspGluIleMetal	500
Q	y	1534	AATCAAGGCAACATGTGCACAGCTACACCTCGCTCAGCTGACGCTGCCCTGCCCTGCAGAA	1593
D	b	500	alleAsnglyLysIleValThrAspPrrrThrLeuAlaGluAlaGluAlaLeuGluLy	520
Q	y	1594	GGCTTGAATACAGGGGGGAGCTGGATCGACCTTGTGGTTCGGCTGCCCCCAAGGA	1653
D	b	520	salaTrpAsnGlnGlyLysPrpileAspLeuValValAlaValCysProProLysGL	540
Q	y	1654	GTATGACGATGACCTGACCTTC	1675
D	b	540	utyrAspAspGluLeuThrPhe	547

RESULT 2

RESUL	Q91XDL	PRELIMINARY;	PRT;	548 AA.
ID	Q91XDL			
AC	G91XDL;			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	RIKEN cDNA 2010016F01 gene.			
OS	2010016F01RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus			
OX	NCBI_Taxid=10090;			
	[1]			
RN	SEQUENCE FROM N.A.			
RP				
RC	TISSUE=COLON;			
RA	Strausberg R.;			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC010819; AAH10819.1; -.			

DR	MGI:1919338; 2010016F0IRik.
DR	InterPro; IPR001478; PDZ.
DR	Pfam; PF00595; PDZ; 3.
DR	PROSITE; PS01016; PDZ; 3.
SQ	SEQUENCE 548 AA; 61711 MW; B00C4F00920ED76B CRC64;
Alignment Scores:	
Pred. No.: .	Length: 548
Score: 2554.50	Matches: 502
Percent Similarity: 93.98%	Conservative: 13
Best Local Similarity: 91.61%	Mismatches: 11
Query Match: 64.04%	Indels: 22
DB: 11	Gaps: 1
US-09-502-945-5 (1-2162) x Q91XD1 (1-548)	
QY	97 ATGCACCGAAGCTGGCCCGGAGATTCCTGGCATAAAGTGGATTTCGTGTTGAATAATGAT 156
Db	1 MetAspArgLysValAlaArGLuPheArgHisLysVaAspPheLeuIleGLuAsnAsp 20
QY	157 GCAGAGAAGCACTACTCTATGATGTCTCGGAATGTACCACAGACCATGACGCTGGCC 216
Db	21 AlaGLuLysAspTyrLeuTyrAspValLeuArgMetTyrHisGLnThrMetAspValAla 40
QY	217 GTGCTCTGGGAGACCTGAAGCTGGTGATCAATGAACCCAGCCGCTGCTCTGTTTGTAT 276
Db	41 ValLeuValGLyAspLeuLysLeuValIleAsnGLuProAsnArgLeuProLeuPheAsp 60
QY	277 GCCATTCGCCGCTGATCCACATCAAGCACAGCTGGAATATGATCAGCTGACCCGCCGG 336
Db	61 AlaIleArgProIleIleProLeuLysHisGLnValGLuTyrAspGLnLeuThrProArg 80
QY	337 CGCTCCAGGAAGCTGAAGGAGGTGCGCTCTGGACCGTCTGCACCCCAGAGCCCTCGCCCTG 396
Db	81 ArgSerArgLysLeuLysGLuValArgLeuAspArgLeuHisProGLuGLyLeuGLyLeu 100
QY	397 AGTGTGCGTGGTCCCTGGAGTTTGCTGPGGCTCTTCATCTCCACCTCATCAAAGCC 456
Db	101 SerValArgGLyGLyLeuGLuPheGLyCysGLyLeuPheIleSerHisLeuIleLysGLy 120
QY	457 GGTCCAGCACACGCTCGGGCTCCAGGTAGGGAGGAGATCGTCCGGATCAATGGATAT 516
Db	121 GLyGLnAlaAspSerValGLyLeuGLnValIGlyAspGLuIleValArgIleAsnGLyTyr 140
QY	517 TCCATCTCCCTCTGATCCCATGAGGAGGTCAACAACCTCATTCGAACCAAGAAAACTGTG 576
Db	141 SerIleSerCysThrHisGLuGLuValIleAsnLeuIleArgThrLysLysThrVal 160
QY	577 TCCATCAAGTAGACACATCGCCCTGTATCCCGTGAAAAGTCTCTCTGATGAGCCCCCTC 636
Db	161 SerIleLysValArgHisIleGLyLeuIleProValLysSerSerProGLuGLuSerLeu 180
QY	637 ACTTGGCAGTATGTGGATCAGTTTGTGCGGAATCTGGGGCGTCCGAGGCAGCCCTGGCC 696
Db	181 LysTrpGLnTyrValAspGLnPheValSerGLuSerGLyGLyValArgGLyGLyLeuGLy 200
QY	697 TCCCTCGGAATCGGAAAAACAAGGAGAAGAAGTCTTCATCAGCTGTGTAGCTCCCGGA 756
Db	201 SerProGLyAsnArgThrThrLysGLuLysValPheIleSerLeuValGLySerArg 220
QY	757 GGCTTGGCTGCAGCATTTCCACGGCCGCCATCCAGAACCTGGGATCTTTATCAGCCAT 816
Db	221 GlyLeuGLyCysSerIleSerSerGLyProIleGLnLysProGLyIlePheValSerHis 240
QY	817 GTGAACCTGGCTCCCTGTCTGCTGAGGTGGGATGGAGATAGGGGACCAGGATTCCTCAA 876
Db	241 ValLysProGLySerLeuSerAlaGLuValGLyLeuGLuThrGLyAspGLnIleValGLu 260
QY	877 GTCAATGGCTCGCACTTCTTAACCTGGATCAAGGAGGCGTGTAAATCTGCTGAAAAAT 936
Db	261 ValAsnGLyIleAspPheThrAsnLeuAspHisLysGLuAlaValAsnValLeuLysSer 280
QY	937 AGCCGCGCCTGACCATCTCCAATTGTAGCTGCGCCGGGAGCGTGTTCATGACACAC 996

RT	the inner ear sensory hair cells, underlies usher syndrome type 1C.",
RL	Natl. Genet. 26:51-55(2000).
DR	EMBL; AF228925; AAG12458.1; .
DR	HSSP; P29476; IQAV.
DR	MGD; MGI:1919338; 2010016F01R1k.
DR	InterPro; IPR001478; PDZ.
DR	InterPro; IPR002965; P_rich_extensn.
DR	Pfam; PF00595; PDZ_3
DR	PRINTS; PR01217; PRICHEXTENS.
DR	SMART; SM00228; PDZ_3.
DR	PROSITE; PS50106; PDZ_3.
DR	SEQUENCE 910 AA; 102284 MW; 427B97953BA5D941 CRC64.
SQ	

Alignment Scores:

Pred. No.:	5,37e-166	Length:	910
Score:	2405.50	Matches:	524
Percent Similarity:	58.06%	Conservative:	20
Best Local Similarity:	55.92%	Mismatches:	39
Query Match:	60.30%	Indels:	354
DB:	11	Gaps:	7

US-09-502-945-5 (1-2162) x Q9ES64 (1-910)

QY	97	ATGACGCGAAAGTGGCCCGAGAAATCCCGCATAAAGTGGATTCTCTGATTGAAATGAT	156
Db	1	MetAspArgLysValAlaArgGluPheArgHisLysValAspPheLeuIleGluAsnAsp	20
QY	157	GCACGAGAGGACTATCTCTATGATGTGCTCGGAATGTACACAGACCATGACGTGGCG	216
Db	21	AlaGluLysAspTyrLeuTyrAspValLeuArgMetTyrHisGlnThrMetAspValAla	40
QY	217	GTGTCGTGGGAGACCTGAAGCTGGTCATCAATGAACCCAGCCGCTGCCTCTGTTTGTAT	276
Db	41	ValLeuValGlyAspLeuLysLeuValIleAsnGluProAsnArgLeuProLeuPheAsp	60
QY	277	GCATTCGGCGCGTGTATCCCACTGAAGCACAGGTGGAATATATCATCAGCTGACCCCGG	336
Db	61	AlaIleArgProLeuIleProLeuLysHisGlnValGluTyrAspGlnLeuThrProArg	80
QY	337	CGTCCAGGAGCTGAAGGAGTGGCTGTGGACCGTCTGCACCCCGAAGGCCCTCGGCCTG	396
Db	81	ArgSerArgLysLeuLysGluValArgLeuAspArgLeuHisProGluGlyLeuGlyLeu	100
QY	397	AGTGTGGCTGGTGGCGAGTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGGC	456
Db	101	SerValArgGlyGlyLeuGluPheGlyCysGlyLeuPheIleSerHisLeuIleLysGly	120
QY	457	GGTCAGCGACAGCGTCGGGGCTCCAGGTAGGGGACGAGATCGTCGGGATCAATGGATAT	516
Db	121	GlyGlnAlaAspSerValGlyLeuGlnValGlyAspGluIleValArgIleAsnGlyTyr	140
QY	517	TCATCTCCTCTGTACCCATGAGGAGGTATCAACCTCATTCGAACCAAGAAAACTGTG	576
Db	141	SerIleSerSerCysThrHisGluGluValIleAsnLeuIleArgThrLysLysThrVal	160
QY	577	TCATCAAGTGAGACACATCGCGCTGTATCCCGCTGAAAAGCTCTCTGTATGAGCCCTC	636
Db	161	SerIleLysValArgHisIleGlyLeuIleProValLysSerSerProGluGlySerLeu	180
QY	637	ACTTGGCAGTATGTGATCAGTTTGTTCGGAATCTGGGGCGTCGGAGGCAGCTGGGC	696
Db	181	LysTrpGlnTyrValAspGlnPheValSerGluSerGlyGlyValArgGlyGlyLeuGly	200
QY	697	TCCCTCGGAAATCGGGAACCAAGAGAGAGGTCTTCATCAGCGCTGTAGGTCCCGA	756
Db	201	SerProGlyAsnArgThrThrLysGluLysLysValPheIleSerLeuValGlySerArg	220
QY	757	GGCCTTGGCTGCAGCATTTCCAGCGGCCCATCCAGAAGCCCTGGCATCTTTATCAGCAT	816
Db	221	GlyLeuGlyCysSerIleSerSerGlyProIleGlnLysProGlyIlePheValSerHis	240
QY	817	GTCAAACTTGGCTCCCTGCTCTGCTGAGTGGGATGGAGATAGGGGACACGATTGTCAA	876

Db	241	VallYsProGlySerLeuSerAlaGluValGlyLeuGluThrGlyAspGlnIleValGlu	260
QY	877	GTCAATGGCGTCGACTTCTCTAACTGGATCAACAAGGAGCTGTAATCTGCTGAATAAT	936
Db	261	VallAsnGlyIleAspPheThrAsnLeuAspHisLysGluIleValAsnValLeuLysSer	280
QY	937	AGCGCAGCTGACCATCTCCATTGTAGCTGAGCTGGCCGGAGCTGTTTCATGACAGAC	996
Db	281	SerArgSerLeuThrIleSerIleValAlaGlyAlaGlyArgGluLeuPheMetThrAsp	300
QY	997	CGGAGCGCTGGCAGAGCGCGCAGCTGAGCTGCACGCCGAGGAGCTTCTCATCGAG	1056
Db	301	ArgGluArgLeuGluGluAlaArgGlnArgGluLeuGlnArgGlnGluLeuMetGln	320
QY	1057	AAGCGCTGGCGATGGAGTCCAAAGATCTCCAGGAGCAGCAGGAGATGGAGCGCAA	1116
Db	321	LysArgLeuAlaMetGluSerAsnLysIleLeuGlnGluGlnGluMetGluArgGln	340
QY	1117	AGGAGAAAGAAATGGCCAGAGAGGCAGAGGAAATGAGAGATACCCGAGAGAGATG	1176
Db	341	ArgArgLysGluIleAlaGlnLysAlaAlaGluGluAsnGluArgTyrArgLysGluMet	360
QY	1177	GAACAGATTGTAGAGGAGGAAGAAGTTTAAAGAAGCAATGGGAAGAGACTGGGGCTCA	1236
Db	361	GluGlnIleSerGluGluGluLysPheLysLysGlnTrpGluGluAspTrpGlySer	380
QY	1237	AAGAACAGCTACTCTTGGCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGC	1296
Db	381	LysGluGlnLeuIleLeuProLysThrIleThrAlaGluValHisProValProLeuArg	400
QY	1297	AAGCCA-----	1302
Db	401	LysProLysSerPheGlyTrpPheTyrArgTyrAspGlyLysPheProThrIleArgLys	420
QY	1302	-----	1302
Db	421	LysAlaLysGluLysLysLysAlaLysTyrAspSerLeuGlnAspLeuArgLysAsnLys	440
QY	1302	-----	1302
Db	441	LysGluLeuGluPheGluGlnLysLeuTyrLysGluLysGluGluMetLeuGluLysGlu	460
QY	1302	-----	1302
Db	461	LysGlnLeuLysIleAsnArgLeuAlaGlnGluValSerGluThrGluArgGluAspLeu	480
QY	1302	-----	1302
Db	481	GluGluSerGluLysThrGlnTyrTrpValGluArgLeuCysGlnThrArgLeuGluGln	500
QY	1302	-----	1302
Db	501	IleSerSerAlaGluAsnGluIleProGluMetThrThrGlyProProProProPro	520
QY	1302	-----	1302
Db	521	SerValSerProLeuAlaProProLeuArgArgPheAlaGlyGlyIleHisLeuHisThr	540
QY	1302	-----	1302
Db	541	ThrAspLeuAspAspIleProLeuAspMetPheTyrTyrProProLysThrProSerAla	560
QY	1302	-----	1302
Db	561	LeuProValMetProHisProProSerValAsnSerProSerLysValProAlaProPro	580
QY	1302	-----	1302
Db	581	ValLeuProSerSerGlyHisValSerSerSerSerProTrpValGlnArgThrPro	600
QY	1302	-----	1302
Db	601	ProProIleProIleProProProProSerIleProThrGlnAspLeuThrProThrArg	620


```
QY 1302 ----- 1302
Db 621 ProLeuProSerAlaLeuGluAlaLeuGlyAsnHisProPheArgThrGlyAspPro 640
QY 1302 ----- 1302
Db 641 GlyHisProAlaAspAspTrpGluAlaAsnThrHisSerGlyLysProSerSerPro 660
QY 1302 ----- 1302
Db 661 ThrThrGluArgSerPheProAlaProLysThrPheCysProSerProGlnProPro 680
QY 1302 ----- 1302
Db 681 ArgGlyProGlyValSerThrIleSerLysProValMetValHisGlnGluHisAsnPhe 700
QY 1302 ----- 1302
Db 701 ValTyrArgProAlaValLysSerGluValLeuProGlnGluMetLeuLysArgMetVal 720
QY 1303 -----AAGTGATTCGGGAATATCATCAGGAAGCTTTGACCCCTA 1341
Db 721 ValTyrGlnThrAlaPheArgGln-AspPheArgLysTyrGluGlyPheAspProTy 740
QY 1342 CTCATGTTCACCCAGACAGATCATCGGGAAGGATGTCGGCTCTACCATCAAGAA 1401
Db 740 rSerMetPheSerProGluGlnIleAlaGlyLysAspValArgLeuLeuArgIleLysLy 760
QY 1402 GGAGGGATCTTAGACCTGGCCCGGAGGCGGTGTGGACTCCGCCATTTGGGAAGTGGT 1461
Db 760 sGluGlySerLeuAspLeuAlaLeuGluGlyGlyValAspSerProValGlyLysValVa 780
QY 1462 CGTTCTCTGTGTATGATGCGGGAGCGTGTGAGCGGATGTCGATGTGGAAGGGGA 1521
Db 780 lValSerAlaValTyrGluGlyGlyAlaAlaGluArgHisGlyGlyValValLysGlyAs 800
QY 1522 CGAGATCATGCAATCAACGGCAAGATGTGTGACAGACTACACCTCGCTGAGGCTGACGC 1581
Db 800 pGluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaGluAl 820
QY 1582 TGCCCTGAGAGGCGCTGATCATCAGGGCGGGGACTGGATGACCTGTGTTGGTGGCTGTG 1641
Db 820 aAlaLeuGlnLysAlaTrpAsnGlnGlyGlyAspTrpIleAspLeuValAlaValIcy 840
QY 1642 CCCCACCAAGGATGTAGTACAGTACCTGACCTTCTGTGAAGTCCAAAAGGGGAACCA 1701
Db 840 sProProLysGluTyrAspAspGluLeuSerSerSerLeuProSerSerAlaAlaGluSer-- 859
QY 1702 AATTCACGCGTTAGGAAACAGTACGCTCGGCCCCACCTCGTGAACACAAAGCCTCGGAC 1761
Db 860 -----ProGlnLeuAla-----ArgLy 865
QY 1762 CAGCCTTGAGAGGCGCCACATCACACACACAGATGGATCTTTGGGACCTCAATCTATC 1821
Db 865 sGlnLeuGlu---AlaTyrGluProValCysArg---HisGlyPhePheLeuGlnLeuGl 883
QY 1822 ACCCAGGAATCTCAAACTCCCTTTGGCCCTGAACACAGGCGCCAGATAAGAACAGCTCGGG 1881
Db 883 uProThrAsnLeuLeuLeu-----LysSerArgGluArgAsnGlnThr-- 897
QY 1882 CCACCTTTTGAAGCCAAATGTGGAGAAAGGAGCAGCCAGCCG 1926
Db 898 -----Asp--ProSerTrpArgProAlaSerSerAlaPro 908
RESULT 5
Q9CVG7
ID Q9CVG7 PRELIMINARY; PRT; 112 AA.
AC Q9CVG7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 2010016F01Rik protein (Fragment).
```

```
GN 2010016F01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK008274; BAB25568.1; -
DR HSSP; Q12923; 3PDZ; 2010016F01Rik.
DR MGD; MGI:1919338; 2010016F01Rik.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER
SQ SEQUENCE 112 AA; 11965 MW; BE42F80C7B3AB196 CRC64;

Alignment Scores:
Pred. No.: 1.6e-31 Length: 112
Score: 548.00 Matches: 105
Percent Similarity: 98.20% Conservative: 4
Best Local Similarity: 94.59% Mismatches: 2
Query Match: 13.74% Indels: 0
DB: 11 Gaps: 0

US-09-502-945-5 (1-2162) x Q9CVG7 (1-112)
QY 1343 TCATGTTTCCACCCAGCAGATCATGGGAAGGATGTCGGCTCTACCATCAAGAAG 1402
Db 1 SerMetPheSerProGluGlnIleAlaGlyLysAspValArgLeuLeuArgIleLysLys 20
QY 1403 GAGGATCTCTTAGACCTGGCCCTGGAAGCGGTGTGGACTCCCGCATTTGGGAAGGTGTC 1462
Db 21 GluGlySerLeuAspLeuAlaLeuGluGlyValAspSerProValGlyLysValVal 40
QY 1463 GTTCTCTGCTGTATGAGCGGGAGCTGCTGAGCGCATGTCGTCATTTGTAAGGGGAC 1522
Db 41 ValSerAlaValTyrGluGlyGlyAlaAlaGluArgHisGlyGlyValValLysGlyAsp 60
QY 1523 GAGATCATGCAATCAACCGCAGATTTGTACAGACTACACCTCGCTGAGGCTGACGCT 1582
Db 61 GluIleMetAlaIleAsnGlyLysIleValThrAspTyrThrLeuAlaGluAlaGluAla 80
QY 1583 GCCCTGCAGAGGCTCGAATCAGGCGGGGACTGGGACTGACCTTGTGGTGGCTGCTGC 1642
Db 81 AlaLeuGlnLysAlaTrpAsnGlnGlyAspTrpIleAspLeuValAlaValLys 100
QY 1643 CCCCAGAGGAGTATGACGATGAGCTGACCTTC 1675
Db 101 ProProLysGluTyrAspAspGluLeuThrPhe 111
```


RESULT 6
 Q9W443 PRELIMINARY: PRT; 493 AA.
 ID Q9W443;
 AC Q9W443;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CG5921 protein.
 DE CG5921.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003436; AAF46116.1; -;
 DR HSSP: Q12959; IPDR.
 DR FlyBase: FBgn0029835; CG5921.
 DR InterPro: IPRO01478; PDZ.
 DR Pfam: PF00595; PDZ; 2.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS0106; PDZ; 2.
 SQ SEQUENCE 493 AA; 50574 MW; C46711C9F19797A CRC64;
 Alignment Scores:
 Pred. No.: 8, 06e-22 Length: 493
 Score: 418.00 Matches: 128
 Percent Similarity: 45.89% Conservative: 62
 Best Local Similarity: 30.92% Mismatches: 157
 Query Match: 10.48% Indels: 67
 DB: 5 Gaps: 12


```
QY 1303 AAGTATTTCGGAATATGAGGAGCT---TTGACCCCTACT----- 1343
Db 406 -----SerThrThrMetLysArgSerHisLeuArgProValAsnSerAlaGlySerG 423
QY 1344 -----CTATGTTTCACCCAGACAGATCATGGGAAGG-----ATGTCGGCTCCT 1389
Db 423 lYlleGlyLeuSerSerGlySerAlaGlySerAlaGlySerAlaGlySerSerGlySerG 443
QY 1390 ACGCATCAAGAGGAGGATCTTATAGACTGGCCCT 1425
Db 443 lYSer-ArgSerGlyGlyValIleAlaProAlaPro 454
RESULT 7
O8WVS2
ID Q8WVS2 PRELIMINARY; PRT; 906 AA.
AC Q8WVS2;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical 96.4 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARY;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014524; AAH14524.1;
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS50106; PDZ; 3.
KW Hypothetical protein.
SQ SEQUENCE 906 AA; 96428 MW; D8D75D87A41A5E05 CRO64;

Alignment Scores:
Pred. No.: 9, 11e-21 Length: 906
Score: 405.00 Matches: 196
Percent Similarity: 36.79% Conservative: 102
Best Local Similarity: 24.20% Mismatches: 285
Query Match: 10.15% Indels: 227
DB: 4 Gaps: 30

US-09-502-945-5 (1-2162) x Q8WVS2 (1-906)
QY 7 CCGGTCCGGTCCGGCTCTTTCCAGCTCTGCGACGCCGGCCACCCGGAAGGACGGTCTG 66
Db 4 ProLeuAspGlyLeuSerValSerSer-----SerThrGlySerLeuGlySer 20
QY 67 TGCAACGACGACGCTGGACCTGGCCCGCCAGCCAGCCGGAAGTGGCCCGGAGAAATTCGG 126
Db 21 AlaAlaGlyAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 40
QY 127 -----CATAGGTGGATTTCTGATGAAATGATGATGATGATGATGATGATGATGATGAT 180
Db 41 GlnLeuHisGlnAlaLeuThrAlaLeuLeuSerGluAlaGluArgGluGlnPheThrHis 60
QY 181 GTCTCGGAATGTACCAACGACCATGGAGCTGGCGTCTGCTGGGAGACCTGGAAGCTG 240
Db 61 CysLeuAsnAlaTyrHisAlaArgArgAsnValPheAspLeuValArgThrLeuArgVal 80
QY 241 GTCATCAATGAACCCAGCCCTGCTGCTCTGTTGATGATGATGATGATGATGATGATGATGAT 300
Db 81 LeuLeuAspSerProValLysArgArgLeuLeuProMetLeuArgLeuValIleProArg 100
QY 301 AACCAACGCTGGAATATGATCAG-----CTGACC 330
Db 101 SerAspGlnLeuLeuPheAspGlnTyrThrAlaGluGlyLeuTyrLeuProAlaThr 120
QY 331 CCGCGGCGC-----TCCAGGAAGCTGAAGGAGGTGCGTCTG 366
```

```
Db 121 ProTyrArgGlnProAlaTyrGlyGlyProAspSerAlaGlyProGlyGluValArgLeu 140
QY 367 GACCGTCTGCAC-----CCGGAAGGCTCGGCTGAGTGTGCTGGTGGCTG 414
Db 141 ValSerLeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySer 160
QY 415 GAGTTTGGCTGTGGCTCTTCATCTCCACCTCATCAAGGCGGTTCAGGACAGACGCTC 474
Db 161 GluHisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGlu 180
QY 475 GGGCTCCAGGTAGGGACAGATGCTCCGATCAATGATATCCATCTCTCTCTGATCC 534
Db 181 GlyLeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThr 200
QY 535 CATGAGGAGGTCAACCACTCATTCGAACCAAGAAACTGTGTCCATCAAGTGAAGACAC 594
Db 201 HisAlaGluAlaValLysAlaLeuLysGlySerLysLeuValLeuSerValTyrSer 220
QY 595 ATCGGCTGTATCCCGGTGAAAGCTCTCTGATGAGCCCTCATCTGGCAGTATGTGAT 654
Db 221 AlaGlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAsp 238
QY 655 -----CAGTTGTGTCGGAATCTGGGGCTGGAGGCGAGCTGGGC----- 696
Db 239 ProGlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeu 258
QY 697 -----TCCCTGGAAATCGGGAAGC-----AAGGAGAAG 726
Db 259 ArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLys 278
QY 727 AAGTCTTTCATCAGCCTGGTAGGTCCCGAGGCTTGGCTGCAGCATTTCAGCGGGCCC 786
Db 279 LysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAla 298
QY 787 ATCCAGAAGCTGGCATCTTTATCAGCCATCTGAAACCTGCTCCTCTCTCTGCTGAGTG 846
Db 299 GluTyrGlyLeuGlyIleTyrIleThrValAspProGlySerGluAlaGluGlySer 318
QY 847 GGATTGGAGATAGGGACGACGATGTCGAAGTCAATGCGCTCGACTTCTTAACCTGGAT 906
Db 319 GlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeu 338
QY 907 CACAAGGAGGTGTAAATGTGCTGAAATAGCCGACGCTGACCATCTCCTCTGCTAGCT 966
Db 339 HisAspGluAlaValArgLeuLysSerSerArgHisLeuLeuLeuThr---ValLys 357
QY 967 GCAGCTGGCCGG-----GAGCTGTTCATGACAGACGCGGAGCGGTGCGACAG 1014
Db 358 AspValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSer 377
QY 1015 GCGCGGACGCTGAGCTGCAGCGGAGGAGCTTCTCATGCAAGCGGCTGCGGATGAG 1074
Db 378 SerArgIleArgGluThrMetAlaAsnSerAlaGlyPheLeuGlyAspLeuThrThrGlu 397
QY 1074 ----- 1074
Db 398 GlyIleAsnLysProGlyPheTyrLysGlyProAlaGlySerGlnValThrLeuSerSer 417
QY 1075 -----TCCAAACAGATCTCCAGGAGCAGCAGAGATGGAGCGCAAGGAGAAA 1125
Db 418 LeuGlyAsnGlnThrArgValLeuLeuGluGluGlnAlaArgHisLeuLeuAsnGluGln 437
QY 1126 GAAATTGCCAGAGGACGAGAGAAATGAGAGATACCGG----- 1167
Db 438 GluHisAlaThrMetAlaTyrTyrLeuAspGlu---TyrArgGlyGlySerValSerVal 456
QY 1168 -----AAGGAGATGGACACAGATTTCTAGAGGAGGAGAGAAATTTAAGAACAA 1215
Db 457 GluAlaLeuValMetAlaLeuPheLysLeuLeuAsnThrHisAlaLysPheSerLeuLeu 476
QY 1216 TGGGAA-----GAAGACTGGGCTCAAGGACACAGCTACTCTTG 1254
```


Qy	535	CATGAGAGGATCATCAACCTCATTTGCAACCAAGAAAACCTGTGTCATCAAAAGTGAGACAC	594
Db	257	HisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSer	276
Qy	595	ATCGGCTGATCCCGTGAAGAGCTCTCTCATGAGCCCTCACTTCGGCAGTATGTGGAT	654
Db	277	AlaGlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrP-----ValAsp	294
Qy	655	-----CAGTTTGTTCGGAATCTGGGGCGCTGGAGCAGCCTGGGC-----	696
Db	295	ProGlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeu	314
Qy	697	-----TCCCTCGGAATCCGGAAAC-----AAGGAGAAG	726
Db	315	ArgGlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyAspGluLys	334
Qy	727	AAGGTCTTCATCAGCTGCTAGGCTCCGAGGCGCTTGCTGCAGCATTTCCACGGCCGCC	786
Db	335	LysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAla	354
Qy	787	ATCCAGAAGCCTGGCATCTTTATCAGCCATGTGAAACCTGGCTCCCTGTCTGTAGGTTG	846
Db	355	GlnTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySer	374
Qy	847	GGATTGGAGATAGGGACAGATTGCGAAGTCAATGCGTCGACTTCTCTACCTGGAT	906
Db	375	GlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeu	394
Qy	907	CACAAGAGAGCTCTAAATGTGCTGAAAATACCGCAGCCCTGACCATCTCCATTGTAGCT	966
Db	395	HisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr---ValLys	413
Qy	967	CGAGCTGCGCCGG-----GAGCTGTTTCATGACAGACCCGGGAGCGGTGGCAGAG	1014
Db	414	AspValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSer	433
Qy	1015	CGCGGGCAGGTGAGCTGCAGCGCAGGAGCTTCATGTCAGAGAGCGGTGGCGATGGAG	1074
Db	434	SerArgIleArgGluThrMetAlaAsnSerAlaGlyPheLeuGlyAspLeuThrThrGlu	453
Qy	1074	-----	1074
Db	454	GlyIleAsnLysProGlyPheTyrLysGlyProAlaGlySerGlnValThrLeuSerSer	473
Qy	1075	-----TCCAACAAGATPCTCCAGGACGACGAGAGATGGAGCGCAAGAGAGAAA	1125
Db	474	LeuGlyAsnGlnThrArgValLeuLeuGluGluGlnAlaArgHisLeuLeuAsnGluGln	493
Qy	1126	GAATTCGCCAAGAGCGACAGAGGAAATGAGAGATACCG-----	1167
Db	494	GluHisAlaThrMetAlaTyrTyrLeuAspGlu---TyrArgGlyGlySerValSerVal	512
Qy	1168	-----AAGGAGATGGAACAGATTGTAGAGGAGGAAGAGAAGTTTAAAGAACAA	1215
Db	513	GluAlaLeuValMetAlaLeuPheLysLeuLeuAsnThrHisAlaLysPheSerLeuLeu	532
Qy	1216	TGGGA-----GAAGATGGGCTCAAAAGGAACACTACTCTTG	1254
Db	533	SerGluValArgGlyThrIleSerProGlnAspLeuGluArgPheAspHisLeuValLeu	552
Qy	1255	CCTAAACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATTCGG	1314
Db	553	ArgArg-----GluIleGluSerMetLysAlaArgGlnPro-----	564
Qy	1315	GAATATGAGGAGGCTTTGACCCCTACTCTATGTTTCACCCAGACAGATCATGGGAA	1374
Db	564	-----	564
Qy	1375	GGATGTCGGGCTCCTACGCATCAAGAAGGAGGATCCTTTAGACCTGGCCCTGGAGCGG	1434
Db	565	-----ProGlyProGly-AlaGlu	570
Qy	1435	T-----GTGGACTCCCCCATTTGG-----	1453

[illegible]


```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ADIPOSE TISSUE;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Oktani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026862; BAB15577.1; -.
DR EMBL; BC015692; AAH15692.1; -.
DR HSSP; Q12923; 3PDZ.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hypothetical protein.
SQ SEQUENCE 517 AA; 55676 MW; 06C9C614283478F7 CRC64;
Alignment Scores:
Pred. No.: 1.2e-11 Length: 517
Score: 278.00 Matches: 78
Percent Similarity: 50.40% Conservative: 48
Best Local Similarity: 31.20% Mismatches: 110
Query Match: 6.97% Indels: 14
DB: 4 Gaps: 6
US-09-502-945-5 (1-2162) x Q9H5P4 (1-517)
QY 349 CTCAAGAGGTGGCTGCTGACCGTCTGCACCCCGAAGGCGCTGGCTGAGTGTGGTGT 408
Db 84 ILeHiserValArgValGluGlySerProAlaGlyArgLeuGlyPheSerValArgGly 103
QY 409 GGCTGAGTTGGCTGTGGCTCTTCATCTCCACCTCAACAAAGCGGTGAGGAGAC 468
Db 104 GlySerGluHisGlyLeuGlyPheValSerLysValGluGlySerSerAlaGlu 123
QY 469 AGCGTGGGCTCAGTAGGGGAGAGATCGTCCGATCAATGGATATTCATCTCTCC 528
Db 124 ArgAlaGlyLeuCysValGlyAspLysIleThrGluValAsnGlyLeuSerLeuGluSer 143
QY 529 TGTACCCATGAGGAGGTCAATCAACCTCATTCGAACCAAGAAACTGTGTCATCAAGTG 588
Db 144 ThrThrMetGlySerAlaValLysValLeuThrSerSerArgLeuHisMetMetVal 163
QY 589 AGACATCGGCTGATCCCGTGAAGAGTCTCTCTGATGAGCCCTCACTGCGAGTAT 648
Db 164 ArgArgMetGlyArgValProGlyIleLysPheSerLysGlyLysThrTrp----- 181
QY 649 GTGGAT-----CAGTTGTGTCGGAATCTCGGGCGCTGCGAGCACCTGGGC 696
Db 182 ValAspValValAsnArgArgLeuValValGluLys-----CysGlySerThrPro 198
QY 697 TCCCTGGAATCGGAAAAAGGAGAAAGTCTTCATCATCGCTGGTAGCTCCCGA 756
Db 199 SerAspThrSerSerGluAspGlyValArgArgIleValHisLeuThrThrSerAsp 218
QY 757 GGC-----CTTGGCTGACATTTCCAGCGGCCCTCCAGAAAGCTGTCATCTTATC 810
Db 219 AspPheCysLeuGlyPheAsnIleArgGlyLysGluPheGlyLeuGlyIleThrVal 238
QY 811 AGCATGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 870
Db 239 SerLysValAspHisGlyGlyLeuAlaGluAsnGlyIleLysValGlyAspGlnVal 258
```

```
QY 871 GTCAAGTCAATGGCTGCTGCTTCTTAACCTGGATCACAAGGAGGCTGTAAATGCTG 930
Db 259 LeuAlaAlaAsnGlyValArgPheAspPheHisSerGlnAlaValGluValLeu 278
QY 931 AAAAATAGCGCGCTGACCATCTCCATCTGACGTGCGCGGAGGTGTTTCATG 990
Db 279 LysGlyGlnThrHisIleMetLeuThrIle---LysGluThrGlyArg-----TyrPro 295
QY 991 ACAGACCGGAGCGCTGGCAGAGCGCGCGAGCTGAGCTGAGCGAGGAGGCTTCTC 1050
Db 296 AlaTyrLysGluMetValSerGluTyrCysTrpLeuAspArgLeuSerAsnGlyValLeu 315
QY 1051 ATGCAGAGCGGCTGGCGATGGAGTCCCAAC 1080
Db 316 GlnGlnLeuSerProAlaSerGluSerSer 325
RESULT 11
Q8VIII PRELIMINARY; PRT; 1665 AA.
AC Q8VIII;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PDZ-domain protein scribble.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6; TISSUE=CEREBELLUM;
RA Mattock K.L., Kurschner C.;
RT "Molecular cloning of mouse Scribble cDNA.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF441233; AAL32489.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003591; LRR_typ.
DR Pfam; PF00560; LRR; 13.
DR Pfam; PF00595; PDZ; 4.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00369; LRR_TYP; 13.
DR SMART; SM00228; PDZ; 4.
DR PROSITE; PS0106; PDZ; 4.
SQ SEQUENCE 1665 AA; 179749 MW; BE1272F4ECBF010E CRC64;
Alignment Scores:
Pred. No.: 2.95e-11 Length: 1665
Score: 275.50 Matches: 128
Percent Similarity: 40.76% Conservative: 66
Best Local Similarity: 26.89% Mismatches: 140
Query Match: 6.91% Indels: 142
DB: 11 Gaps: 24
US-09-502-945-5 (1-2162) x Q8VIII (1-1665)
QY 271 TTTGATGCATTCGCGCTGATCCCATGACAGCAGGTGAATATGATCAGCTGACC 330
Db 695 PheAspGlnAlaAsnAsnLeuIleGluProAlaArgIleGluGluGluLeuThr 714
QY 331 CCCCAGCGCTCCAGGAAGCTGAAGAGGTGCTGAGCCGCTGCCACCCCGAAGGCTC 390
Db 715 -----LeuThrIleValArg-----GlnThrGlyLeu 724
QY 391 GGCCTGAGTGTGCGTGGTGGCTGGAGTTGGCTGT----- 426
Db 725 GlyIleSerIleAlaGlyLys-----GlySerThrProTyrLysGlyAspGlu 742
QY 427 GGCCTCTTCATCTCCACCTCATCAAGCGGTGAGCAGCAGCAGCGTGGGCTCCAGTA 486
Db 743 GlyIlePheIleSerArgValSerGluGlyProAlaAlaArgAlaGlyValArgVal 762
QY 487 GGGGAGGAGATCGTCGGGATCAATGAATATTCATCTCTCTCTCTCTCTCTCTCTCT 546
```



```
Db 763 GlyAspLysLeuLeuGluValAsnGlyValAlaLeuGlnAspAlaGluHisGluAla 782
QY 547 ATCAACCTCAATCGA---ACCAAGAAACTGTGCTCCATCAAGTG----- 588
Db 783 ValGluAlaLeuArgGlyAlaGlyAlaValGlnMetArgValTrpArgGluArgMet 802
QY 589 -----AGACATATGGCTGATCCCGCTGAAAGCTCTCCTGAT---GAGCCC 633
Db 803 ValGluProGluAsnAlaValThrIleThrProLeuArgProGluAspAspTyrSerPro 822
QY 634 CTCACCTGGCAGTATGTGGATCAGTTGTGCGGAATCTGGGCGCGCGCAGCAGCTG 693
Db 823 ArgGluTrpArg-----GlyGlyGlyLeuArgLeuProLeu 834
QY 694 GGCTCCCTCGAAATCGGGAACAAGAGAGAGAGTCTTCATCAGCCTGGTAGGCTCC 753
Db 835 LeuGlnProGluThrProValSerLeuArgGlnArgHisAlaAlaCysLeuValArgSer 854
QY 754 ---CGAGCCTTGCTGCTGACGATTTCCAGCGCGGCCCATCCAGAGCCT----- 798
Db 855 GluLysGlyLeuGlyPheSerIleAlaGlyLysGlySerThrProTyrArgAlaGly 874
QY 799 -----GGCATCTTATCAGCCATGTCGAAGTGGCGCTGCTGCTGCTGAGTGGGA--- 849
Db 875 AspGlyIlePheIleSerArgIleAlaGluGlyAlaAlaHisArgAlaGlyThr 894
QY 850 TTGGAGATAGGACGACAGATGTGCAAGTGCAGTGGCTGCTGCTGCTGCTGCTGCTG 909
Db 895 LeuGlnValGlyAspArgValLeuSerIleAsnGlyValAspMetThrGluAlaArgHis 914
QY 910 AAGAGCCTTAATGTGCTGGAATAATAGCCGACGCTGACCATCTCCATCTGAGTGC 969
Db 915 AspHisAlaValSerLeuLeu---ThrAlaAlaSerProThrIleSerLeuLeuGlu 933
QY 970 GCTGGCGGGAGCTGTCATGACAGACCGGCGGCTGCGCAGAGGCGGCGAGCTGAG 1029
Db 933 uArg-----GluThrGly-----GlyThrTyrPr 941
QY 1030 CTCAGCGCGCAGGAGCTTCTCATGCGAGAGCGGCTGCGGAGTGGATCCCAAGATCCCT 1089
Db 941 oProSerProProHisSer-----SerProThrProAlaAl 954
QY 1090 CAGGACGACGAGAGATGGAGCGG-----AAAGAGAAAGAAATGGCCAGAGGCA 1143
Db 954 aThrValAlaAlaThrValSerThrAlaValProGlyGluProLeuProArg----- 972
QY 1144 GCAGAGAAATAGAGATACCGGAGGAGATGGAACAGATTTAGAGGAGGAGAGAAG 1203
Db 972 ----- 972
QY 1204 TTTAAGACAATGGGAAGAAGACTGGGCTCAAAGAACAGCTACTCTTGCTTAAACCC 1263
Db 973 -----LeuSerPr 975
QY 1264 ATCAGTGTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATTTCGGAAATATGA 1323
Db 975 oSerLeuLeuAlaThrAlaLeu----- 982
QY 1324 GGAGGCTTTGACCCCTACTCTATGTTCAACCCAGCAGATCATGGGAAGGATGTCGG 1383
Db 983 -GluGly-----ProTyrProVal-----GluGluIleCys 992
QY 1384 GCTCTCTACGCATCAAGAAGGAGGATCCCTAGACCTGGCCCTGGGAAGCGGTGTGAC--- 1441
Db 992 sLeuProArg-----AlaGlyGlyProLeuGlyLeuSerIleValGlyGlySerAspHis 1010
QY 1442 -----TCCCCCATTTGG-----AAGGTGGTGGCTTTCTGCTGTGTATGAGCG 1482
Db 1010 sSerSerHisProPheGlyValGlnAspProGlyValPheIleSerLysValLeuProAr 1030
QY 1483 GGGAGCTGCTGAGCGGATGTGGCATTTGTGAAGGGGACGAGATCATGGCAATCAAGG 1542
Db 1483 ----- 1542
```

```
Db 1030 gGlyLeuAlaAlaArg---CysGlyLeuArgValGlyAspArgIleLeuAlaValAsnGlu 1049
QY 1543 CAGATTGTGACAGACTACACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1588
Db 1049 yGlnAspValArgGluAlaThrHisGlnGluAlaValSerAlaLeu 1064
RESULT 12
075970 PRELIMINARY; PRT; 2042 AA.
ID 075970;
AC 075970;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Multi PDZ domain protein MUPP1.
GN MUPP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Eng L., Kravitsky G., Clapham D.E.;
RT "Human homolog of MUPP1 protein."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF093419; RAC61870.1;
DR HSSP; Q12959; IPDR.
DR InterPro; IPR004172; L27.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF02828; L27; 1.
DR Pfam; PF00595; PDZ; 13.
DR SMART; SM00228; PDZ; 13.
DR PROSITE; PS0106; PDZ; 13.
SQ SEQUENCE 2042 AA; 218595 MW; F0E1DD029749A762 CRC64;
Alignment Scores:
Pred. No.: 4.88e-11 Length: 2042
Score: 273.00 Matches: 146
Percent Similarity: 36.78% Conservative: 71
Best Local Similarity: 24.75% Mismatches: 170
Query Match: 6.84% Indels: 203
DB: 4 Gaps: 25
US-09-502-945-5 (1-2162) x 075970 (1-2042)
QY 11 TCGCGGTCCGGCTCTTTCCAGCTCTGCGACGCGGCGACCGAA-----GGAACG 61
Db 1585 SerGlnSerLeuMetValProGlnSerGlySerProGluProGluSerIleArgAsnThr 1604
QY 62 GGTCTGTCACACGACGCTGCGACCTGCGCCAGCCATGACCGAAAGTGGCCCGAGAT 121
Db 1605 SerArgSerSerThrProAlaIlePheAla-Ser---AspProAlaThrCysProIle 1623
QY 122 TCCGCGATAAGTGGATTTTCTGATTGAAATGATGC----- 158
Db 1623 eProGly-----CysGluThrThrIleGluIleSerLys 1634
QY 159 -----AGAGAAGGACTATCTATGATGCTGTG----- 186
Db 1634 sclyArgThrGlyLeuGlyLeu-SerIleValGlyGlySerAspThrLeuLeuGlyAla 1654
QY 187 -----CGATGTACCCAG-----ACCATGACGTCGCGCTGCTGCTG 226
Db 1654 leIleIleHisGluValTyrGluGlyAlaAlaCysLysAspGlyArgLeuTrpAlaG 1674
QY 227 GAGACTGAAGCTGCTCATCAAT----- 249
Db 1674 lyAspGlnIleLeuGluValAsnGlyIleAspLeuArgLysAlaThrHisAspGluAla 1694
QY 250 -----GAACCCAGCTCTGCTCTGTTGATGCCATTCGCGCGCTGA 292
Db 1694 leAsnValLeuArgGlnThrProGlnArgValArgLeu---ThrLeuTyrArgAspGlu 1713
```



```
QY 293 TCCCACTGAAGCAGCAGGTGGATATGATCAGCTGACCCCGCGCTCCAGGAAGCTGA 352
Db 1713 laProTyLysGluGluGluValCysAspThrLeuThr-----
QY 353 AGGAGGTGCGTCTGACCGCTGTCACCCGAGCCCTGGCGCTGAGTGTGGTGGGCC 412
Db 1726 -----IleGluLeuGlnLysLysProGlyLysGlyLeuGlyLeuSerIleValGlyLysA 1744
QY 413 TGGAGTTGGCTGGGCTCTTCATCTCCACCTCATCAAGGGCGGTGAGCAGACGCG 472
Db 1744 rgAsn---AspThrGlyValPheValSerAspIleValLysGlyGlyIleAlaAspAlaA 1763
QY 473 TCGGG---CTCAGGTAGGAGCAGAGATCTCCGGATCAATGGATPATTCCATCTCTCT 529
Db 1763 spGlyArgLeuMetGlnGlyAspGlnIleLeuMetValAsnGlyGluValArgAsnA 1783
QY 530 GTACCCATGAGGAGGTATCAACCTCATTCGA---ACCAAGAAACTGTGTCCATCAAG 586
Db 1783 laThrGlnGluAlaValAlaAlaLeuLeuLysCysSerLeuGlyThrValThrLeuGluV 1803
QY 587 TCAGACACATCGCCTGATCCCGTGAAAGCTCTCCTGATGAGCCCTCACTTGGCAGT 646
Db 1803 al-----GlyArgIleLysAlaGlySerSerThrSerGluSerLeu----- 1816
QY 647 ATGTGGATCAGTTTGTTCGGAATCTGGGGCGCTGCGAGGCGACCTGGGCTCC-----C 700
Db 1817 -----GluSerSerSerLysLysAsnAlaLeuAlaSerGluIleG 1830
QY 701 CTGGAATCGGAAACAAAGAGAGAGAGGTCTTCATCAGCCTGGTAGCTCCCGAGGCC 760
Db 1830 lngLeuArgThrValGluMetLysLys-----GlyProThrAspSerL 1845
QY 761 TTGGCTGACAGTTTCCACCGGC-----CCCATCCAGAGCCTGCATCTTTATCA 811
Db 1845 euGlyIleSerIleAlaGlyValGlyValGlySerProLeuGlyAspValProIlePheIleA 1865
QY 812 GCATCTGAAACCTGCTCCCTCTGCTGCTGAGTG---GGATTGGAGATAGGGGACCA 868
Db 1865 laMetMetHisProThrGlyValAlaAlaGlnThrGlnLysLeuArgValGlyAspArgI 1885
QY 869 TTGTCAAGTCAATGGCTGCTGCTGCTCTTAACCTGATCACAAGAGGCTGTAATGTGC 928
Db 1885 leValThrIleCysGlyThrSerThrGluGlyMetThrHisThrGlnAlaValAsnLeuL 1905
QY 929 TGAATAATAGCCGACGCTGACCATCTCCATCTGATGCTGCGCGGGAGCTTTTCA 988
Db 1905 euLysAsnAla---SerGlySerIleGluMetGlnValValAlaGly-GlyAspValSer 1923
QY 989 TGACAGCGGGAGCGGCTGGCAGAGCGCGCAGCGCTGAGTGCAGCGGAGGAGCTTC 1048
Db 1924 ValValThrGly----- 1927
QY 1049 TCATGCAGAGCGGCTGGCGATGGATCCCAAGATCCTCCAGGAGCAGCAGGATGG 1108
Db 1927 ----- 1927
QY 1109 AGCGGCAAGGAGAGAAATTTGCCAGAGAGCAGCAGAGGAAATGAGAGATACCGGA 1168
Db 1927 ----- 1927
QY 1169 AGGAGATGGAACAGATTGTAGAGGAGAGAGAAAGTTTAAGAAAGCAATGGGAAGAAG 1228
Db 1927 ----- 1927
QY 1229 GGGCTCAAGGAAACAGCTACTCTTGCCTAAACCATCACTGCTGAGGTACACCCAGTAC 1288
Db 1928 -----HisGlnGlnGlu 1931
QY 1289 CCTTCGCAAGCAAGTATTTCCGGAATATGAGGAGGCTTGTACCCCTACTCTATG 1348
Db 1932 ProLaserSerSerLeuSerPhe-----ThrGlyLeuThrSerSerSerIle 1947
QY 1349 TTCACC-----CCAGAGCAGATCATGGGAAGGATGTCCGGCTCTTACGCATC 1396
```

```
Db 1948 PheGlnAspAspLeuGlyProGlnCys-----LysSerIleThrLeuGluArg--- 1964
QY 1397 AAGAGGAGGATCCTTAGACCTGGCCCTGGAAGGCGTGGACTCCCCCATTTGGG--- 1453
Db 1965 ---GlyProAspGlyLeuGlyPheSerIleValGlyIleValGlySerProHisGlyAsp 1983
QY 1454 ---AAGGTGGTCTGCTCTGCTATGAGCGGGAGCTGCTGAGCGGCATGTGGCATT 1510
Db 1984 LeuProIleThrValLysThrValPheAlaLysGlyAlaLaserGluAspGlyArgLeu 2003
QY 1511 GTAAAGGAGGAGATCATGCGCAATCAAGCGCAAGATTGTGACAGACTACACCTGGCT 1570
Db 2004 LysArgGlyAspGlnIleIleAlaValAsnGlyGlnSerLeuGluGlyValThrHisGlu 2023
QY 1571 GAGCTGACGCTGCCCTGCAGAAG 1594
Db 2024 GluAlaValAlaIleLeuLysArg 2031
RESULT 13
Q9NTP3 PRELIMINARY; PRT; 202 AA.
ID Q9NTP3 AC Q9NTP3; DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BA108L7.3 (Novel PDZ (DHR, GLCF) domain protein) (Fragment).
GN BA108L7.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Heath P.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133215; CAB92801.1; -.
DR HSSP; P29476; 1QAV.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
FT NON_TER 1 202
SQ NON_TER 202 202
SQ SEQUENCE 202 AA; 21709 MW; 506507836F8845C CRC64;
Alignment Scores:
Pred. No.: 2, 22e-11 Length: 202
Score: 272.00 Matches: 69
Percent Similarity: 52.74% Conservative: 37
Best Local Similarity: 34.33% Mismatches: 83
Query Match: 6.82% Indels: 12
DB: 4 Gaps: 5
US-09-502-945-5 (1-2162) x Q9NTP3 (1-202)
QY 379 CCCGAAGGC---CTCGGCTGAGTGTGGTGGCTGGCTGGAGTTGGCTGGGCTCTTC 435
Db 6 ProAlaGlyArgLeuGlyPheSerValArgGlyGlySerGluHisGlyLeuGlyIlePhe 25
QY 436 ATCTCCACCTCATCAAGGCGGTGAGCAGACAGCGTCCGGCTCCAGGTAGGCGACGAG 495
Db 26 ValSerLysValGluGlySerSerAlaGluArgAlaGlyLeuCysValGlyAspLys 45
QY 496 ATCTCGCGGATCAATGGATATTCTCTCTCTTACCATGAGGAGGTCTATCAACCTC 555
Db 46 IleThrGluValAsnGlyLeuSerLeuGluSerThrMetGlySerAlaValLysVal 65
QY 556 ATTCGAACCAAGAACTGTGTCCATCAAGAGTGAGACACATCGGCTGTATCCCGTGA 615
Db 66 LeuThrSerSerArgLeuHisMetMetValArgMetGlyArgValProGlyIle 85
QY 616 AGCTCTCTGTAGAGCCCTCACTTGGCAGTATGTGGAT-----CAGTTTGTG 663
```


Db 86 LysPheSerLysGluLysThrThrTrp-----ValAspValValAsnArgArgLeuVal 103
 QY 664 TCGGAATCTGGGGCGTGGAGGAGCCCTGGCTCCCTCGGAAATCGGAAACAGAGAG 723
 Db 104 ValGluLys-----CysGlySerThrProSerAspThrSerSerGluAspGlyVal 120
 QY 724 AAGAGGTCTTCATCAGCCTGGTAGCTCCCGAGGC-----CTTGGCTGCAGCATTTCC 777
 Db 121 ArgArgIleValHisLeuThrThrThrSerAspPheCysLeuGlyPheAsnIleArg 140
 QY 778 AGCGCCCATCCAGAACCTGGCATCTTATCAGCCATGTGAACCTGGCTCCCTGTCT 837
 Db 141 GlyGlyLysGluPheGlyLeuGlyIleThrValSerLysValAspHisGlyGlyLeuAla 160
 QY 838 GCTAGGTGGGATTGGACATAGGACCATGTCGAAGTCAATGGCGTGCAGCTTCTCT 897
 Db 161 GluGluAsnGlyIleLysValGlyAspGlnValLeuAlaAlaAsnGlyValArgPheAsp 180
 QY 898 AACCTGGATCAGAGGCTGTAATCTGCTGTAATAATAGCCGAGCCTGACCATCTCC 957
 Db 181 AspIleSerHisSerGlnAlaValGluValLeuLysGlyGlnThrHisIleMetLeuThr 200
 QY 958 ATT 960
 Db 201 Ile 201

RESULT 14

Q96C69 ID Q96C69 PRELIMINARY; PRT; 835 AA.
 AC Q96C69;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 87.7 kDa protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC014632; AAH14632.1;
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 3.
 DR PROSITE; PS0106; PDZ; 3.
 KW Hypothetical protein.
 FT NON_TER 1 1
 SQ SEQUENCE 835 AA; 87669 MW; F59DF419D7AAF991 CRC64;

Alignment Scores:

Pred. No.: 4,33e-11 Length: 835
 Score: 271.50 Matches: 178
 Percent Similarity: 32.13% Conservative: 90
 Best Local Similarity: 21.34% Mismatches: 252
 Query Match: 6.81% Indels: 316
 DB: 4 Gaps: 31

US-09-502-945-5 (1-2162) x Q96C69 (1-835)

QY 289 CTGATCCACTGAAGCACCAGGTGGATATGATCAGCTGACCCCGCGCTCCAGGAAG 348
 Db 30 IleThrProLeuArgProGluAspAspThrSerPro-----ArgGluArgArg 45
 QY 349 CTGAAGAGGTGGCTGTGGACCGTCTCTCACCCCGAA-----384
 Db 46 GlyGlyGlyLeuArgLeuProLeuLeuProGluSerProGlyProLeuArgGlnArg 65
 QY 385 -----GGCTCGCGCTGAGTGTGGTGGC---411
 Db 66 HisValAlaCysLeuAlaArgSerGluArgGlyLeuGlyPheSerIleAlaGlyGlyLys 85

QY 412 -----CTGGAGTTTGGCTGGGGCTCTTCTATCTCCACCTCATCAAGGC 456
 Db 86 GlySerThrProThrArgAlaGlyAspAlaGlyIlePheValSerArgIleAlaGluGly 105
 QY 457 GGTGAGGAGCAGACGCTCGG-----CTCAGGTAGGAGCAGAGATCGTCCGATCAATGGA 513
 Db 106 GlyAlaAlaHisArgAlaGlyThrLeuGlnValGlyAspArgValLeuSerIleAsnGly 125
 QY 514 TATTCATCTCTCTCTTACCCATGAGGAGGTCTCAACCTCATTT---CGAACCAAGAAA 570
 Db 126 ValAspValThrGluAlaArgHisAspHisAlaValSerLeuLeuThrAlaAlaSerPro 145
 QY 571 ACTGTGTCATCAAGTGCAGACATCGGCTGTATCCCGGTGAAGACTCTCTCT 624
 Db 146 ThrIleAlaLeuLeuGluArgGluAlaGlyGlyProLeuProProSerProLeuPro 165
 QY 625 -----GATGAGCCCTCACTTGGCAGTATCTGTCAGTCTGTCGGAATCTGGGGC 678
 Db 166 HisSerSerProThrAlaAlaValAlaThrThrSerIleThrAlaThrProGly 185
 QY 679 GTGCGAGC-----687
 Db 186 ValProGlyLeuProSerLeuAlaProSerLeuLeuAlaAlaLeuGluGlyProTyr 205
 QY 688 -----AGCTGGGCTCCCTCGAAATCGGAAACACAGGAG 723
 Db 206 ProValGluGluIleArgLeuProArgAlaGlyPro-----218
 QY 724 AAGAAGCTTTCATCAGCTGTAGCTCCCGAGGCTTGGCTCCAGCATTTCCAGCGC 783
 Db 219 -----LeuGlyLeuSerIleValGly-----GlySerAspHisSerSerHis 232
 QY 784 CCC-----ATCCAGAACCTGGCATCTTATCAGCCATGTGAACCTGCTCTCTCT 837
 Db 233 ProPheGlyValGlnGluProGlyValPheIleSerLysValLeuProArgGlyLeuAla 252
 QY 838 GCTGAGTGGGATTGGAGATAGGGACCATGTCGAAGTCAATGCGTGCAGTCTCTCT 897
 Db 253 AlaArgSerGlyLeuArgValGlyAspArgIleLeuAlaValAsnGlyGlnAspValArg 272
 QY 898 AACCTGGATCACAAGGAGGCTGTAATGTCTG-----930
 Db 273 AspAlaThrHisGlnGluAlaValSerAlaLeuLeuArgProCysLeuGluLeuSerLeu 292
 QY 930 -----930
 Db 293 LeuValArgArgAspProAlaProProGlyLeuArgGluLeuCysIleGlnLysAlaPro 312
 QY 931 -----AAAAATAGCCGC 942
 Db 313 GlyGluArgLeuGlyIleSerIleArgGlyGlyAlaArgGlyHisAlaGlyAsnProArg 332
 QY 943 AGCCTGACC-----ATCTCCATTGTAGCTGCAGTGCCTCGC 978
 Db 333 AspProThrAspGluGlyIlePheIleSerLysValSerProThrGlyAlaAlaGlyArg 352
 QY 979 GAGCTGTTTCATGACACCGGAGCGGCTGGCAGAGCGCGGCGGCGTGCAGTGTG-----1032
 Db 353 AspGlyArgLeuArgValGlyLeuArgLeuGluValAsnGlnSerLeuLeuGly 372
 QY 1033 -----CAGCGGAGGAGCTTTCATGCA-----1055
 Db 373 LeuThrHisGlyGluAlaValGlnLeuLeuLeuLeuSerValGlyAspThrLeuThrValLeu 392
 QY 1056 -----GAAGCGCTGCGGATGGAGTCCACAGATCCTCCAGGA-----1094
 Db 393 ValCysAspGlyPheGluAlaSerThrAspAlaLeuGluValSerProGlyValIle 412
 QY 1095 -----GCAGCAGGAGATGGA-----CGGCAAG 1118
 Db 413 AlaAsnProPheAlaAlaGlyIleGlyHisArgAsnSerLeuLeuSerIleSerIle 432

Db 835 GlyAlaAlaHisArgAlaGlyThrLeuGlnValGlyAspArgValLeuSerIleAsnGly 854
QY 514 TATTCATCTCTCTCTACCCATGAGGAGGTCAACACCTCATTT---CGAACCAAGAA 570
Db 855 ValAspValThrGluAlaArgHisAspHisAlaValSerLeuLeuThrAlaAlaSerPro 874
QY 571 ACTGTGTCATCAAGTGAACATCGCCCTGATCCCGTGAAAGCTCTCTCT 624
Db 875 ThrIleAlaLeuLeuGluAlaGlyGlyProLeuProProSerProLeuPro 894
QY 625 -----GATGAGCCCTCCTGAGTGGCATCTGTGTCGGAATCTGGGGC 678
Db 895 HisSerSerProThrAlaAlaValAlaThrThrIleThrAlaThrProGly 914
QY 679 GTGCGAGGC----- 687
Db 915 ValProGlyLeuProSerLeuAlaProSerLeuLeuAlaAlaAlaLeuGluGlyProTyr 934
QY 688 -----AGCCTGGGCTCCCTGGAAATCGGGAACAAAGAG 723
Db 935 ProValGluGluIleArgLeuProArgAlaGlyGlyPro----- 947
QY 724 AAGAGGCTTCATCAGCCTGGTAGGCTCCGAGCCCTGGCTGCAGCATTTCCAGCGGC 783
Db 948 -----LeuGlyLeuSerIleValGly-----GlySerAspHisSerHis 961
QY 784 CCC-----ATCCAGAGCCGTCATCTTATCAGCCATGTGAACCTGGCTCTCTCT 837
Db 962 ProPheGlyValGlnGluProGlyValPheIleSerLysValLeuProArgGlyLeuAla 981
QY 838 GCTGAGTGGGATTGGAGTGGGACACAGATTTCGAAGTCAATGGCTGACTTCTCT 897
Db 982 AlaArgSerGlyLeuArgValGlyAspArgIleLeuAlaValAsnGlyGlnAspValArg 1001
QY 898 AACCTGGATCACAGGAGGCTGTAATGTGCTG----- 930
Db 1002 AspAlaThrHisGlnGluAlaValSerAlaLeuLeuArgProCysLeuGluLeuSerLeu 1021
QY 930 ----- 930
Db 1022 LeuValArgArgAspProAlaProProGlyLeuArgGluLeuCysIleGlnLysAlaPro 1041
QY 931 -----AAAAATAGCCGC 942
Db 1042 GlyGluArgLeuGlyIleSerIleArgGlyAlaArgGlyHisAlaGlyAsnProArg 1061
QY 943 AGCCTGACC-----ATCTCCATTGAGCTGAGCTGCGCGG 978
Db 1062 AspProThrAspGluGlyIlePheIleSerLysValSerProThrGlyAlaAlaGlyArg 1081
QY 979 GAGCTGTTCATGACAGCCGCGGAGGCTGCGAGAGCGCGGCGCTGAGCTG----- 1032
Db 1082 AspGlyArgLeuArgValGlyLeuArgLeuLeuGluValAsnGlnGlnSerLeuLeuGly 1101
QY 1033 -----CAGCGCCAGGAGCTTCTCATGCA----- 1055
Db 1102 LeuThrHisGlyGluAlaValGlnLeuLeuArgSerValGlyAspThrLeuThrValLeu 1121
QY 1056 -----GAAGCGCTGGCGATGGAGTCCAAACAGATCCTCCAGGA----- 1094
Db 1122 ValCysAspGlyPheGluAlaSerThrAspAlaAlaLeuGluValSerProGlyValIle 1141
QY 1095 -----GCAGCAGAGATGGA-----GCGCAAG 1118
Db 1142 AlaAsnProPheAlaAlaGlyIleGlyHisArgAsnSerLeuGluSerIleSerIle 1161
QY 1119 GAGAAAGAAATTCGCCAGAGAGCAGACAGAGGAAATGAGATACCGAAGGAGATGGA 1178
Db 1162 AspArgGluLeuSerProGluGlyProGlyLysGluLysGluLeuPro-----Gly 1178
QY 1179 ACAGATTGTAGAGGAGAGAGAGTTTAAGCAATGGGAAGA----- 1223
Db 1179 GlnThrLeuHisTrpGlyProGluAlaThrGluAlaAlaGlyArgGlyLeuGlnProLeu 1198

QY 1224 AGACTGGGCTCAAGGAACAGCTACTCTTCCCTAAAACCATCACTGCTGAGGTACACCC 1283
Db 1199 LysLeuAspTyrArgAlaLeuAlaAlaVal-----Pro 1209
QY 1284 AGTACCCTTCGCAAGCAAGTGTATTCGGAATATGAGGAAGGCTTTCACCCCTACT 1343
Db 1210 SerAlaGlySerValGlnArgValProSerGlyAlaAlaGlyLysMetAlaGluSer 1229
QY 1344 CTATGTTCAACCCAGACAGCATCATGGGAAGGATGTCGGCTCTACGCATCAAGAAGG 1403
Db 1230 ProCysSerProSerGlyGln-GlnProProSerProSerProAspGluLeuProAl 1249
QY 1404 AGGATCCTTAGA----- 1416
Db 1249 aAsnValLysGlnAlaTyrArgAlaPheAlaAlaValProThrSerHisProGluAs 1269
QY 1417 -----CCTGGCCCTGGNAG----- 1431
Db 1269 pAlaProAlaGlnProProThrProGlyProAlaAlaSerProGluGlnLeuSerPheAr 1289
QY 1432 -----CGCTGTGACTCCCCCATTTGGGAAGT-----GG 1460
Db 1289 gGluArgGlnLysTyrPheGluLeuGluValArgValProGlnAlaGluGlyProProLy 1309
QY 1461 TCGTTTCTGCTGTATGACGGGGAGCTCTGAGCGGCATGCTGGCATTTGTAAAGGGG 1520
Db 1309 sArgValSerLeuVal----- 1315
QY 1521 ACNAGATCATCGGCATCAACGCGAAGATTGTGACAGACTACACCCCTGGCTGAGCGTGACG 1580
Db 1315 yAlaAspAspLeuArgLysMetGlnGlu----- 1324
QY 1581 CTGCGCTGCAGAAGCGCTGGAATCAGGGCGGAGCTGGATCGACCTTGTGGTTGCCGTCT 1640
Db 1325 -----GluGluAlaArgLysLeuGlnGln----- 1332
QY 1641 GCCCCCAAGGAGTATGACGATGACCTGCTTCTGTGAAGTCCAAAGGGGAAACC 1700
Db 1333 -----LysArgAlaGlnMetLeuArgGluAlaAlaGlyAl 1346
QY 1701 AAATTCACGCGTTAGGAACACAGTGAGCTCGGCGCCACCTCTGTAACACAAACCTCGGA 1760
Db 1346 aGluAlaArgLeuAlaLeuAspGlyGluThrLeuGlyGluGluGlnGluAspGluG 1366
QY 1761 CCAGCCTTGAGAGAGCCACATGACACACAGATGGCATCTCTTGGGACCT----- 1812
Db 1366 nProProTrpAlaSerProSerProThrSerArgGlnSerProAlaSerProProProLe 1386
QY 1813 -----GAATCTATCACCAGGAATCTCAAC 1838
Db 1386 uGlyGlyGlyAlaProValArgThrAlaLysAlaGluArgArgHisGlnGluArgLeuAr 1406
QY 1839 TCCTTTTGGCCCTGAACAGGCGCCAGATAGGGAACAGCTCGGGCCACTTTT-TTGAAGCG 1897
Db 1406 gValGlnSerProGluProAlaProGluArgAlaLeuSerProAlaGluLeuArgAl 1426
QY 1898 CAATGTGGGAAAGGAGGAGCCAGCCCTTTGGGAGAAGATCTCAAGGATCCAGACTCT 1957
Db 1426 a-----LeuGluAlaGluLysArgAlaLeuTrp---ArgAlaAlaArgMetLysSerLe 1443
QY 1958 CATTCCTTCTCTCTGCCCCAGTGAATTTGCTCTCTCCAGCTTTGG-----GGGAC 2008
Db 1443 uGluGlnAspAlaLeuArgAlaGlnMetValLeuSerArgSerGlnGluGlyArgGlyTh 1463
QY 2009 T-----CCTTCCTTGAACCCCTAATAAGAC 2032
Db 1463 rArgGlyProLeuGluArgLeuAlaGluAlaProSerProAlaProThrProSerProTh 1483
QY 2033 CCCACTGGAGTCTCTC----- 2048
Db 1483 rProValGluAspLeuGlyProGlnThrSerThrSerProGlyArgLeuSerProAspPh 1503

Tue Mar 25 09:28:42 2003

Qy 2049 -----TCTCTCCATCCCTCTCTGCCC 2072
||||| ||||| |||
Db 1503 eAlaGluGluLeuArgSerIeuGluProSerProSerPro 1516

Search completed: March 21, 2003, 12:59:57
Job time : 103.143 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 11:21:37 ; Search time 27.1901 Seconds
(without alignments)
8801.668 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346

Sequence: 1 ggaattcctctgttcgaagt.....aaacaaaagttaaaattt 2885

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xl
-Q/cgn2.1/USPTO_spool/US09502945/runat_14032003.101058.19100/app_query.fasta_1.10979
-DB=SwissProt 40 -OPMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pio -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09502945@cgn.1.113@runat_14032003.101058.19100 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	4793	89.7	1122	1	HD45_HUMAN
2	4565.5	85.4	1113	1	HD45_MOUSE
3	2732.5	51.1	1080	1	HD41_CHICK
4	2691	50.3	1084	1	HD44_HUMAN
5	2579	48.2	1011	1	HD49_HUMAN
6	950.5	17.8	588	1	HD49_MOUSE
7	771.5	14.4	1215	1	HD46_HUMAN
8	771	14.4	1149	1	HD46_MOUSE
9	705.5	13.2	687	1	HD41_SCHPO
10	672	12.6	706	1	HD41_YEAST
11	651.5	12.2	798	1	YLFN_CAEEL
12	490.5	9.2	359	1	Y130_ARCFU
13	385	7.2	331	1	Y894_METHH
14	358.5	6.7	304	1	Y453_SYN3
15	348	6.5	343	1	Y535_METJA
16	345	6.5	385	1	ACUC_STAXY
17	331.5	6.2	387	1	ACUC_BACSU
18	316.5	5.9	480	1	HD41_CHICK

19	312	5.8	461	1	HD41_CAEEL	017695 caenorhabdi
20	309	5.8	488	1	HD42_MOUSE	P70288 mus musculus
21	306.5	5.7	482	1	HD41_HUMAN	Q13547 homo sapien
22	306.5	5.7	482	1	HD41_MOUSE	Q09106 mus musculus
23	304	5.7	488	1	HD42_HUMAN	Q92769 homo sapien
24	303.5	5.7	433	1	RPD3_YEAST	P32561 saccharomyc
25	303	5.7	480	1	HD11_XENLA	Q91695 xenopus lae
26	301	5.6	452	1	HOS2_YEAST	P53096 saccharomyc
27	297.5	5.6	434	1	PHD1_SCHPO	Q13298 schizosacch
28	297	5.6	480	1	HD12_XENLA	Q42227 xenopus lae
29	289	5.4	576	1	HD41_STRPU	P56518 strongyloce
30	288	5.4	513	1	HDAC_MAIZE	P56521 zea mays
31	285	5.3	520	1	HDAC_DROME	Q94517 drosophila
32	280	5.2	488	1	HD42_CHICK	P56519 gallus gall
33	279.5	5.2	660	1	YHL1_EBV	P03181 epstein-bar
34	268	5.0	507	1	HD42_CAEEL	Q09440 caenorhabdi
35	265.5	5.0	501	1	HDAC_ARATH	O22446 arabidopsis
36	261	4.9	424	1	HD43_MOUSE	O88895 mus musculus
37	260	4.9	428	1	HD43_CHICK	P56520 gallus gall
38	260	4.9	428	1	HD43_HUMAN	O15379 homo sapien
39	255.5	4.8	660	1	YHL1_EBV	P03181 epstein-bar
40	254.5	4.7	1255	1	MUC1_HUMAN	P15941 h mucin 1 p
41	253.5	4.7	2715	1	TRX2_HUMAN	O9um06 homo sapien
42	251	4.7	341	1	APHA_MYCRA	O48935 mycoplasma r
43	242	4.5	2161	1	SHK1_HUMAN	O91566 homo sapien
44	239	4.5	1733	1	VNUA_PRVKA	P33485 pseudorabie
45	234	4.4	1783	1	RAA3_CHLRE	Q9fec4 chlamydomon

ALIGNMENTS

RESULT 1
HD45_HUMAN STANDARD: PRT: 1122 AA.
AC Q9UOL6; O60528; O60340;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone deacetylase 5 (HD5) (Antigen NY-CO-9).
GN HDAC5 OR KIA0600.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9238449; PubMed=10220385;
RA Grozinger C.M., Hassig C.A., Schreiber S.L.;
RT "Three proteins define a class of human histone deacetylases related to yeast Hda1p.";
RL proc. Natl. Acad. Sci. U.S.A. 96:4868-4873(1999).
RN [2]
RP SEQUENCE OF 189-1085 FROM N.A.
RC TISSUE=Colorectal carcinoma;
RA Scanlan M.J., Chen Y., Williamson B., Gure A.O., Stockert E.,
RT "Characterization of human colon cancer antigens recognized by autologous antibodies.";
RL Int. J. Cancer 76:652-658(1998).
RN [3]
RP TISSUE=Brain;
RC MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ohikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RT "Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";
RL DNA Res. 5:31-38(1998).
CC -!- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4). HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL

CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS.
 CC REPRESSIONS TRANSCRIPTION (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
 CC FAMILY, HD SUBFAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF132608; AAD29047.1; -;
 DR EMBL: AF039691; AAC18040.1; -;
 DR EMBL: AB011172; BAA25526.1; -;
 DR Genbank: HGNC:14068; HDAC5..
 DR MIN: 605315; -;
 DR InterPro: IPR000286; His_deacetylase.
 DR Pfam: PF00850; Hist_deacetyl; 1.
 DR PRINTS: PR01270; HDASUPER.
 KW Hydrolase; Nuclear protein; Transcription regulation; Repressor.
 FT DOMAIN 47 52 POLY-GLY.
 FT DOMAIN 85 92 POLY-GLN.
 FT DOMAIN 596 599 POLY-GLU.
 FT DOMAIN 1099 1104 POLY-ALA.
 FT DOMAIN 684 1028 HISTONE DEACETYLASE.
 FT CONFLICT 671 671 S -> N (IN REF. 2).
 FT CONFLICT 684 684 G -> S (IN REF. 2).
 FT CONFLICT 1026 1026 E -> K (IN REF. 2).
 FT CONFLICT 1074 1074 E -> G (IN REF. 2).
 FT CONFLICT 1085 1085 T -> M (IN REF. 2).
 SQ SEQUENCE 1122 AA; 121991 MW; CF4BBB8B9A288FEC CRC64;

Alignment Scores:

Pred. No.: 3 43e-193 Length: 1122
 Score: 4793.00 Matches: 927
 Percent Similarity: 99.57% Conservative: 3
 Best Local Similarity: 99.25% Mismatches: 4
 Query Match: 89.66% Indels: 1
 DB: 1 Gaps: 0

US-09-502-945-2 (1-2885) x HDA5_HUMAN (1-1122)

QY 2 GAATTCCTCTTCGAAGTCAAGAGGCCACACGAGGGGCTCAACATTCCTCCCA 61
 Db 189 GluPheLeuLeuSerLysSerLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 208
 QY 62 CAGCACCCAAATGCTGGGAGCCACCATGCTCTTTGGACACAGATTCCCTCCCGAG 121
 Db 209 GlnHisProLysCysTrpGlyAlaHisHisAlaSerLeuAspGlnSerSerProGln 228
 QY 122 AGCGGCCCCCTGGGAGCGCTCCCTCAACAACTGCCCTTTCCTGGGCGCTTACGACAGT 181
 Db 229 SerGlyProProGlyThrProProSerTyLysLeuProLeuProGlyProTyLysSer 248
 QY 182 CGAGAGAGTTCCTCCCGCAAAACAGCCTCTGACCACTTGAAGTGGCTTCAAGG 241
 Db 249 ArgAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg. 268
 QY 242 CTAACAACAAGGTGCTGAGCGGAGAACAGTCCCTCCTCGCTCGCAAGGATGGACT 301
 Db 269 LeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArgLysAspGlyThr 288
 QY 302 GTTATTAGCACCTTTAAGAAAGAGAGCTGTGATCACAGGTGCCGGGCTGGGGCGTGG 361
 Db 289 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyAlaGlyProGlyAlaSer 308
 QY 362 TCCGTGTGAACAGCCAGCCCGGCTCCGGCCCCAGCTCTCCCAACAGCTCCCAACAGCACC 421
 Db 309 SerValCysAsnSerAlaProGlySerGlyProSerSerProAsnSerSerHisSerThr 328

QY 422 ATCGCTGAGAATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCAG 481
 Db 329 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGln 348
 QY 482 CACCGAGCCCTCCCTCTGGACAGCTCCCAACACAGTTGAGCTCTACAGGTCTCTTCT 541
 Db 349 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTyThrSerProSer 368
 QY 542 CTCCCAACATCTCCCTAGGGCTGACAGCCACCGTCTCCTCACCACCTCACACCTCACT 601
 Db 369 LeuProAsnIleSerLeuGlyLeuGlnAlaThrValThrValThrAsnSerHisLeuThr 388
 QY 602 GCCTCCCGAAGTGTGACACAGCAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 661
 Db 389 AlaSerProLysLeuSerThrGlnGlnAlaGluArgGlnAlaLeuGlnSerLeuArg 408
 QY 662 CAGGTGGCAGCTGACCGCAAGTTCATGAGCACATCTCTATTCTGCTGGCTGCTGCTG 721
 Db 409 GlnGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCysLeuLeu 428
 QY 722 GCGTGGCAGTGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 781
 Db 429 GlyValAlaLeuGluGlyAspGlySerProHisGlyHisAlaSerLeuLeuGlnHisVal 448
 QY 782 CTGTTGCTGGAGCAGCGCCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 841
 Db 449 LeuLeuLeuGluGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 468
 QY 842 TCCCACTAGTACGGGTGAACGTGTGGCCACGACGATGCGGAGGAGGAGGAGGAGGAGG 901
 Db 469 SerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLysLeuPro 488
 QY 902 CGGATCGGCCCTGAGCCGCACTCAGTCTCCGCTGCGGAGGAGTCCCGAGGCGCTG 961
 Db 489 ArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGlnAlaLeu 508
 QY 962 CAGCAGCTGCTCATCAACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1021
 Db 509 GlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnLysGlnGln 528
 QY 1022 CTACAGCTGGCAAGATCTTCACCAAGACAGGGAGGAGTCCCGAGGAGCAGCAGCAGCAG 1081
 Db 529 LeuGlnLeuGlyLysIleLeuThrLysThrGlyLeuLeuProArgGlnProThrHis 548
 QY 1082 CTGAGGAGCAGAGGAGGAGGAGTCAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1141
 Db 549 ProGluGluThrGluGluGluLeuThrGluGlnGlnGluValLeuLeuGlyGlyAla 568
 QY 1142 CTGACCATGCCCCGGGAGGCTCCACAGAGAGTGGAGGAGCAGCAGAGAGCAGCAGGAGG 1201
 Db 569 LeuThrMetProArgGluGlySerThrGluSerGluSerThrGlnGluAspLeuGluGlu 588
 QY 1202 GAG 1261
 Db 589 GluAspGluGluGluAspGlyGluGluGluAspCysIleGlnValLysAspGluGlu 608
 QY 1262 GCGGAGAGTGTGCTGAGGAGGGCGCCGACTTGGAGGAGCCTGCTGCTGGATACAAAAA 1321
 Db 609 GlyGluSerGlyAlaGluGluGlyProAspLeuGluGluProGlyAlaGlyTyLysLys 628
 QY 1322 CTGTTCTCAGATGCCCAACCGCTCAACCTTTGAGGTGTACCAAGCGCCCTCAGCGCTG 1381
 Db 629 LeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyGlnAlaProLeuSerLeu 648
 QY 1382 GCCACTGTGCCCCACCAAGCCCTGGCGGTACCCAACTCTCCCTCCCTGCTGCCCTGGGGGC 1441
 Db 649 AlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaProGlyGly 668
 QY 1442 ATGAAGAACCCCCCAGACCAACCGCTCAAGCACCTCTTCCACCAAGTGTGGTCTAGCAC 1501
 Db 669 MetLysSerProProAspGlnProValLysHisLeuPheThrThrGlyValValTyAsp 688
 QY 1502 ACGTTTCATGATAAGCACCACTGTCATGTGCGGGAACACACAGTGCACCTTGATGCTGCT 1561

Db 689 ThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGluHisAla 708
 QY 1562 GCGCGGATCCAGAGCATCTGTCGCGGCTGCAGGAGACAGCGCTGCTTAGCAAGTGCAG 1621
 Db 709 GlyArgIleGlnSerIleTrpSerArgLeuGlnThrGlyLeuLeuSerLysCysGlu 728
 QY 1622 CGGATCCGAGTCCGAAAGCCACGCTAGATGAGATCCAGACAGTGCACCTCTGAATACCAC 1681
 Db 729 ArgIleArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGluThrHis 748
 QY 1682 ACCCTGCTCTATGGACGACGCTCCCTCAACCGGAGAGAGCTAGACAGCAAGAGTTCCTC 1741
 Db 749 ThrLeuLeuArgGlyThrSerProLeuAsnArgGlnLysLeuAspSerLysLysLeuLeu 768
 QY 1742 GGTCCCATCAGCCAGAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1801
 Db 769 GlyProIleSerGlnLysMetTyrAlaValLeuProCysGlyGlyIleGlyValAspSer 788
 QY 1802 GACACCGTGGATGAGATGCATCTCTCCAGTGTGTCGATGCGCATGGAGTGGGCTGCTG 1861
 Db 789 AspThrValTrpAsnGlnMetHisSerSerAlaValArgMetAlaValIleArg 808
 QY 1862 CTGAGCTGGCTTCAAGGTGGCTGCAGGAGAGCTCAGAGATGATGCTGCTGCTGCTGCTGCT 1921
 Db 809 LeuGluLeuAlaPheLysValAlaAlaGlyGluLeuLysAsnGlyPheAlaIleArg 828
 QY 1922 CCCCCAGGACACCGCGGAGGATCCACAGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1981
 Db 829 ProProGlyHisLysAlaGluSerThrAlaMetGlyPheCysPheAsnSerVal 848
 QY 1982 GCCATCACCAGCAAACTCTACAGCAGAGTGAACGTGGGCAAGTCCCTCATCTGCTGAC 2041
 Db 849 AlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeuIleValAsp 868
 QY 2042 TGGGACATTCACCATGCAATGGGACCCAGCAGCGCTTCAATGATGACCCCTCTGCTGCTC 2101
 Db 869 TrpAspIleHisLysGlyAsnGlyThrGlnGlnAlaPheTyrAsnAspProSerValLeu 888
 QY 2102 TACATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2161
 Db 889 TyrIleSerLeuHisArgTyrAspAsnGlyAsnPhePheProGlySerGlyAlaProGlu 908
 QY 2162 GAGTTGGTGGAGACACGCGGCGGTGCAATGTAACGTGGCATGTCAGCAGGAGTGTG 2221
 Db 909 GluValGlyGlyGlyProGlyValGlyTyrAsnValAsnValAlaTrpThrGlyGlyVal 928
 QY 2222 GACCCCGCATTTGAGACGTGGAGTACCTTACAGCCTTACAGCAGTGGTGTGATGCCCAT 2281
 Db 929 AspProIleGlyAspValGlyThrAlaPheArgThrValValMetProIle 948
 QY 2282 GCGCACGATTCACCTGATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2341
 Db 949 AlaHisGluPheSerProAspValValLeuValSerAlaGlyPheAspAlaValGluGly 968
 QY 2342 CATCTGCTCTCTGGTGGCTACTCTGTCCACCGCCAGATGTTTGGCCACTTGCACAGG 2401
 Db 969 HisLeuSerProLeuGlyGlyTyrSerValThrAlaArgCysPheGlyHisLeuThrArg 988
 QY 2402 CAGCTGATGACCTGGCAGGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2461
 Db 989 GluLeuMetThrLeuAlaGlyArgValValLeuAlaLeuGluGlyGlyHisAspLeu 1008
 QY 2462 ACCGCCATCTGTGATGCTCTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2521
 Db 1009 ThrAlaIleCysAspAlaSerGluAlaCysValSerAlaLeuLeuSerValGluLeuGln 1028
 QY 2522 CCCTTGATGAGCAGCTCTTGCAGCAAAAGCCCAACATCAACGAGTGGCGCAGCTAGAG 2581
 Db 1029 ProLeuAspGluAlaValLeuGlnGlnLysProAsnIleAsnAlaValAlaThrLeuGlu 1048
 QY 2582 AAGTCATCAGATCCAGACCAACACTGGAGCTGTGTGAGAGAGTTCGCGCTGCTGCTGCTGCT 2641

Db 1049 LysValIleGluIleGlnSerLysHisTrpSerCysValGlnLysPheAlaAlaGlyLeu 1068
 QY 2642 GCGCGTCCCTGCGAGGGGCGCCAGCAGTGTGAGCAGCAGGAGCCGAAAT-GTGAAGCC 2700
 Db 1069 GlyArgSerLeuArgGluAlaGlnAlaGlyGluThrGluGluAlaGluThrValSerAla 1088
 QY 2701 ATGGCCCTGTGTGGTGGGCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2760
 Db 1089 MetAlaLeuLeuSerValGlyAlaGluGlnAlaGlnAlaAlaAlaAlaArgGluHisSer 1108
 QY 2761 CCCAGCGCGCAGGAGCCCATGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2822
 Db 1109 ProArgProAlaGluGluProMetGluGlnGluProAlaLeu 1122

RESULT 2
 HDAS_MOUSE
 ID HDAS_MOUSE STANDARD; PRT; 1113 AA.
 AC Q922V6; Q9JL73;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Histone deacetylase 5 (HD5) (Histone deacetylase mHDA1).
 GN HDAC5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Fetal;
 RX MEDLINE=99107904; PubMed=9891014;
 RA Verdel A., Khochbin S.;
 RT "Identification of a new family of higher eukaryotic histone
 deacetylases. Coordinate expression of differentiation-dependent
 chromatin modifiers.";
 RL J. Biol. Chem. 274:2440-2445(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=20107033; PubMed=10640276;
 RA Kao H.Y., Downes M., Ordentlich P., Evans R.M.;
 RT "Isolation of a novel histone deacetylase reveals that class I and
 class II deacetylases promote SMRT-mediated repression.";
 RL Genes Dev. 14:55-66(2000).
 CC -1- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
 THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
 CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
 CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS.
 CC REPRESS TRANSCRIPTION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APAH
 CC FAMILY. HD SUBFAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF006602; RAD09834.2; -;
 DR EMBL; AF207748; AAF31418.1; -;
 DR MGD; MGI:1333784; Hdac5.
 DR InterPro; IPR000286; Hist_deacetylase.
 DR Pfam; PF00850; Hist_deacetyl; 1.
 DR PRINTS; PR01270; HDASUPER.
 KW Hydrolase; Nuclear protein; Transcription regulation; Repressor.
 FT DOMAIN 47 52 POLY-GLY.
 FT DOMAIN 85 92 POLY-GLN.
 FT DOMAIN 577 588 POLY-GLU.
 FT DOMAIN 675 1019 HISTONE DEACETYLASE.
 FT CONFLICT 7 7 S -> SA (IN REF. 2).


```

Db 838 SerValAlaIleThrAlaLysLeuLeuGlnGlnLysLeuSerValGlyLysValLeuIle 857
QY 2036 GTGACTGGACATTCACCATGCAATGCGACCCACCGCGCTTACATGACCCCTCT 2095
Db 858 ValAspTrpAspIleHisGlyAsnGlyThrGlnAlaPheTyrAsnAspProSer 877
QY 2096 GTGCTCTACATCTCTGCTGCTATGACACGGGAACCTCTTCCAGGCTCGGGCT 2155
Db 878 ValLeuTyrIleSerLeuHisArgTyrAspAsnGlyAsnPheProGlySerGlyAla 897
QY 2156 CCTGAAGAGTTGGTGAGACAGCGCGTGGGTAACTGTAACCTGCGATGGACAGAA 2215
Db 898 ProGluGluValGlyGlyProGlyValGlyTyrAsnValAsnValAlaIleThrGly 917
QY 2216 GGTGTGGACCCCTTCCAGCGTGGAGTACCTTACAGCTTCCAGGACAGTGTGATG 2275
Db 918 GlyValAspProIleGlyAspValGluTyrLeuThrAlaPheArgThrValMet 937
QY 2276 CCATTGCCACAGTTCACCTGATGTGCTAGTCTCCGCGCGGTTTGTATGCTGT 2335
Db 938 ProIleAlaGlnGluPheSerProAspValValLeuValSerAlaGlyPheAspAlaVal 957
QY 2336 GAAGACATCTGCTCTGCTGGTGGCTACTGTGTACCGCCAGATGTTTGGCCACTTG 2395
Db 958 GluGlyHisLeuSerProLeuGlyGlyTyrSerValThrAlaArgCysPheGlyHisLeu 977
QY 2396 ACCAGCAGCTGATCACCTCGCGCGGCGCGGGTGGTGTCTGCGCTGGAGGAGCCAT 2455
Db 978 ThrArgGlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGluGlyHis 997
QY 2456 GACTTGACCGCATCTGTGATGCCCTCTGAAGCTGTGTCTCGCTCTGCTCAAGTAAAG 2515
Db 998 AspLeuThrAlaIleCysAspAlaSerGluAlaCysValSerAlaLeuLeuSerValGlu 1017
QY 2516 CTGAGCCCTTGGATGAGCAGCTTGTGACGAAAGCCCAACATCAACGAGTGGCCAG 2575
Db 1018 LeuGlnProLeuAspGluAlaValLeuGlnGlnLysProSerValAsnAlaValAlaThr 1037
QY 2576 CTAGAGAAGTCAATCAGATCCAGACCAACACTGGAGCTGTGTCAGAGTTCCCGCT 2635
Db 1038 LeuGluLysValIleGluIleGlnSerLysHisTrpSerCysValGlnArgPheAlaAla 1057
QY 2636 GGTCTGGCGGCTCTGCTCGAGGGGCCCAAGCAGGCTGAGACCGAAGACCGCAAT-GTG 2694
Db 1058 GlyLeuGlyCysSerLeuArgGluAlaGlnThrGlyGlyLysGluGluAlaGluThrVal 1077
QY 2695 AACGCTATGGCTTCTGCTGTGGTGGGCGGCGAAGCGCCCAAGCTGCGGACCGCGGAA 2754
Db 1078 SerAlaMetAlaLeuLeuSerValGlyAlaGluGlnAlaGlnAlaValAlaThrGlnGlu 1097
QY 2755 CACAGCCCGCGCGCAGAGCGCCATGGAGCAGGAGCGCTGCGCTG 2802
Db 1098 HisSerProArgProAlaGluProMetGluGlnGluProAlaLeu 1113

RESULT 3
HDA4_CHICK
ID HDA4_CHICK STANDARD; PRT: 1080 AA:
AC P83038;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone deacetylase 4 (HD4).
GN HDAC4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Takechi S., Azuma R., Nakayama T.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

```

```

CC CC FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
CC THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS (BY
CC SIMILARITY). REPRESSES TRANSCRIPTION (BY similarity).
CC -|- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -|- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY. HD SUBFAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB052839; BAB60957.1;
CC InterPro: IPR000286; Hist_deacetylase.
CC Pfam: PF00850; Hist_deacetyl; 1.
CC PRINTS: PR01270; HDASUPER.
CC Hydrolase; Nuclear protein; Transcription regulation; Repressor.
CC KW DOMAIN 651 1080 HISTONE DEACETYLASE.
CC SEQUENCE 1080 AA; 119467 MW; B6416E2C43F1428C CRC64;

```

```

Alignment Scores:
Pred. No.: 2,31e-107 Length: 1080
Score: 2732.50 Matches: 576
Percent Similarity: 71.22% Conservativeness: 112
Best Local Similarity: 59.63% Mismatches: 185
Query Match: 51.11% Indels: 94
DB: 1 Gaps: 20

```

US-09-502-945-2 (1-2885) x HDA4_CHICK (1-1080)

```

QY 2 GAATTCCTCTTGTGCAAGCTCAAGAGAGCCACACAGCGCGCTCAACCATTCCTCCCA 61
Db 176 GluPheValLeuAsnLysLysLysAlaLeuAlaHisArgAsnLeuAsnHisCysIleSer 195
QY 62 CAGCACCCCAATGCTGG-----GGAGCCCAACCATGCTCTTTGGACAGAGTCCCT 115
Db 196 SerAspProArgPheTrpTyrGlyLysThrGlnHisSerSerLeuAspGlnSerPro 215
QY 116 CCCAGAGCGCGCCCTGGGACCGCTCCCTCTCAAACTGCTTTCCTGGCGCCCTAC 175
Db 216 ProGlnSerGlyValSerGlyThr-----TyrAsnHisProValLeuGlyMetTyr 232
QY 176 GACAGTCGAGACGACTTCCCTCCGCAAAACAGCTCTGAACCAACTTGAAGTGCCT 235
Db 233 AspSerLysAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysLeuArg 252
QY 236 TCAAGGCTAAACAGAGGTGGCTGAGCGGAGAGCAGTCCCTCTCGCTGCGCAAGAT 295
Db 253 SerArgLeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArgLysAsp 272
QY 296 GGGACTGTTATAGCACCTTTAAGAAGAGAGCTGTTGAGATCACAGGTGCGCGCCTGG 355
Db 273 GlyProValValThrAlaLeuLysLysArgProLeuAspValThr----- 287
QY 356 CGCTGCTCGGTGTAAACAGCGCACCGCGCTCCGCGCCCGCTCCCAACAGCTCCAC 415
Db 288 ---AspSerAlaCysAsnSerAlaProGlySerGlyProSerProAsnAsnSerSer 306
QY 416 AGCACCATC---GCTGAGAATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATG 472
Db 307 AsnAsnIleSerAlaGluAsnGlyIleThrGlySerValThrSerIleGlnAlaGluThr 326
QY 473 CTCCTCAGCACCGAGCCCTCCCTCTCGACAGCTCCCGCCCAACAGTTTCAGCCTTACAG 532
Db 327 SerLeuAlaHisArgLeuValAsnArgGluGlySerValThrGlnLeuProLeuTyrThr 346
QY 533 TCTCTCTCTGCGCCCAACATCTCCCTAGGCTGCGAGCGCCAGGTCACCTGTCCACCACTCA 592

```


Db 347 SerProSerLeuProAsnIleThrLeuGlyLeuProAlaThrGlyProSerSerGlyGly 366
Qy 593 CACCTCACTGCTCCCGAAGCTGTCACACAGCAGAGCGCCGAGAGCGCCCTCCAG 652
Db 367 -----SerAlaGlnAlaGluArgLeuAlaIlePro 378
Qy 653 TCCCTCGGGCAG-----GGTGGCAGCTGACCGCGCAAGTTCATGAGCACA 697
Db 379 AlaLeuGlnGlnArgIleSerLeuPheProGlyThrHisLeuThrProTyrLeuSerThr 398
Qy 698 TCCCTATTCTGCTGCTGCTGGCGTGGCACTGGAGGGCGACGGGAGCCCGCCAGGG 757
Db 399 ThrThr-----LeuGluArgAspGlyGlyThr---Ala 408
Qy 758 CATGCTCCCTGCTGCAGCATGCTGCTGCTGAGCAGAGCGCCGCGACAGACACCCCTC 817
Db 409 HisAsnProLeuGlnHisMetValLeuLeuGlnProThrAlaGlnThrProLeu 428
Qy 818 ATT---GCTGTGCCACTCCACGGCAGTCCCACTAGTACGGGTGAACGTGTGGCCACC 874
Db 429 ValThrGlyLeuProLeuHisAlaGlnSer---LeuValGlyGlyGluArgValSerPro 447
Qy 875 AGCATGGGAGCGGTAGGCAAGCTCCCGGCGATCGGCCCTGAGCGCGACTCATGCTCTCA 934
Db 448 Ser-----IleHisLysLeuArgGlnHisArgProLeuGlyArgThrGlnSerAla 464
Qy 935 CCGCTGCCGACAGTCCCGCAGCGCTGCTGAGTGGTGCATGCAACACACACAGCAGCAG 994
Db 465 ProLeuProGlnAlaAlaGlnAlaLeuGlnLeuValIleGlnGlnGlnHisGlnGln 484
Qy 995 TTCTGGAGAAAGCAGAAG-----CAGCAGCAGCTACACGTGGGCAAGATCCTC 1042
Db 485 PheLeuGluLysHisLysGlnGlnPheGlnGlnGlnGlnLeuHisIleAsnLysIleIle 504
Qy 1043 ACCAAGACAGGGAGCTGCCAGGACGCCACCCACCCCTGAGGAGCAGAGGAGGAG 1102
Db 505 SerLysProAsnGluProAlaArgGlnHisGluSerHisProGluGluThrGluGluGlu 524
Qy 1103 CTGACGGAGCAGCAGAGGTCTGCTGGGGAGGAGCCCTGACCATGCCCGGGAGGGC 1162
Db 525 LeuArgGluHisGlnAlaLeu----- 532
Qy 1163 TCCACAGAGATGAGACACACAGAGACCTGGAGGAGGAGGAGAGAGATGGG 1222
Db 533 -----GluGluProTyrSerAspArgValSerSerGlnLysGluValProGly 548
Qy 1223 GAGGAGGAGGAGGATGATCAGGTTAAGACAGGAGGAGGAGGAGGTGCTGAGGAG 1282
Db 549 -----LeuAlaAsnMetValGlnValLysGlnGluProIleGluSerAspGluGlu 566
Qy 1283 GGGCCGCGACTTGGAGAGCCTGGTCTGGATACAAAAA----- 1321
Db 567 AlaGluProGlnGlnGluLeuGluSerGlyGlnArgGlnAlaGluGlnGluLeuPhe 586
Qy 1322 -----CTGTTCTCAGATCCCAACCGCTGCCAACCTTTCCAGGTGTACCAAGCG 1369
Db 587 ArgGlnGlnAlaLeuLeuGlnGlnArgIleHisGlnLeuArgAsnTyrGlnAla 606
Qy 1370 CCCCTCAGCCTGGCCACTGTGCCC-----CACCAGCCCTGGGCGGTACC 1414
Db 607 SerLeuGluAlaAlaGlyMetProValSerPheGlyHisArgProLeuSerArgAla 626
Qy 1415 CAATCCTCCCTGCTGCC-----CCTGGGGCGATGAAGACCCCGCAGACCAACCC 1465
Db 627 GlnSerSerProAlaSerAlaThrPheProMetSerValGlnGluProThrLysPro 646
Qy 1466 GTCAAGCACCTTCCACACAAAGTGTGTTCTACGACACTTCATGCTAAAGCAGCAGTGC 1525
Db 647 -----ArgPheThrThrGlyLeuValTyrAspThrLeuMetLeuLysHisGlnCys 663
Qy 1526 ATGTGCGGGAACACACATGTCCTGAGCATGTGTCGGCGGATCCAGCATCTGTGTC 1585
Db 664 ThrCysGlyAsnThrAsnSerHisProGluHisAlaGlyArgIleGlnSerIleThrPse 683

Qy 1586 CGGCTCAGGAGACAGGCGCTGCTTAGCAAGTGCAGCGGATCCGAGGTCCGAAAGCCAG 1645
Db 684 ArgLeuGlnGlnThrGlyLeuArgGlyLysCysGluCysIleArgGlyArgLysAlaThr 703
Qy 1646 CTAGATGAGATCCAGACAGCTGCTGAATACCAACACCTCTCTATATGGGACCACTGCC 1705
Db 704 LeuGluGluLeuGlnThrValHisSerGluAlaHisThrLeuLeuTyrGlyThrAsnPro 723
Qy 1706 CTCACCCGCGAGAGCTAGACAGCAAGAAGTGTGCTGCTGCCATCAGCAGCAAGATGAT 1765
Db 724 LeuAsnArgGlnLysLeuAspSerLysLysLeuLeuGlySerLeuThr---SerMetPhe 742
Qy 1766 GCTGTGCTGCTGCTGGGGCATCCGGGTGGACAGTGCACACCGTGTGGAAATGAGATGCAC 1825
Db 743 ValArgLeuProCysGlyGlyValGlyValAspSerAspThrIleTrpAsnGluValHis 762
Qy 1826 TCTCCAGTGTCTGCGCATGTCAGTGGCGTGCCTGCTGAGAGCTGCGCTTCAAGGTGGCT 1885
Db 763 SerSerGlyAlaAlaArgLeuAlaValGlyCysValIleGluLeuValPheLysValAla 782
Qy 1886 GCAGGAGACCTCAAGATGGATTGCTTCATCATCCGCGCCCGCAGACACACCGCGAGGAA 1945
Db 783 ThrGlyGluLeuLysAsnGlyPheAlaValAlaValArgProProGlyHisHisAlaGluGlu 802
Qy 1946 TCCACAGCCATGGGATCTCTTCAACTCTGTAGCCATCACCCGCAAACTCTCTACAG 2005
Db 803 SerThrProMetGlyPheCysTyrPheAsnSerValAlaIleAlaLysLeuLeuGln 822
Qy 2006 CAGAAGTTGAACGTGGGCAAGTCTCATGCTGAGTGGGACATTCACCATGCAATGCG 2065
Db 823 GlnArgLeuAsnValSerLysIleLeuIleValAspTrpAspValHisHisGlyAsnGly 842
Qy 2066 ACCCAGCAGCGTCTACATGACCCCTGCTGCTCTACATCTCTGTCATCTGCTATGAC 2125
Db 843 ThrGlnGlnAlaPheTyrAsnAspProAsnValLeuTyrIleSerLeuHisArgTyrAsp 862
Qy 2126 AAGCGGAACCTCTTTCAGGCTCTGGGCTCCTCAAGAGGTGTGGAGCAGCAGCGGTG 2185
Db 863 AspGlyAsnPhePheProGlySerGlyAlaProAspGluValGlyThrGlyAlaGlyVal 882
Qy 2186 GGGTACATGTGAACGTGGCATGGACAGGAGGTGTGGACCCCGCCCATTTGGAGAGCGGAG 2245
Db 883 GlyPheAsnValAsnMetAlaPheThrGlyGlyLeuAspProMetGlyAspThrGlu 902
Qy 2246 TACCTTACAGCCTTCAGGACAGTGGTGTATGCCATTCACCCACGAGTCTCACCTGATGTG 2305
Db 903 TyrLeuThrAlaPheArgThrValValMetProIleAlaAsnGluPheAlaProAspVal 922
Qy 2306 GTCTAGTCTCCGCGGGTTTGATGTGTGAAGGACATCTGCTCTCTGCTGGTGGCTAC 2365
Db 923 ValLeuValSerSerGlyPheAspAlaValGluGlyHisProThrProLeuGlyGlyTyr 942
Qy 2366 TCTGTACCCCGCAGATGTTTGGCCACTTCACAGGAGGAGTGTGACCCCTGGAGGGGGC 2425
Db 943 AsnLeuSerAlaLysCysPheGlyTyrLeuThrLysGlnLeuMetGlyLeuAlaGlyGly 962
Qy 2426 CGGTGTGTGCTGGCCCTGGAGGAGGCGCATGACTTGACCGCCCATCTGTGATGCTCTGAA 2485
Db 963 ArgValValLeuAlaLeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGlu 982
Qy 2486 GCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2545
Db 983 AlaCysValSerAlaLeuLeuGlyAsnGluLeuAspProLeuLysValLeuGln 1002
Qy 2546 CAAAGCCCAACATCAACGAGTGGCCACGCTAGAGAAAGTCACTGAGATCCGAGCAAA 2605
Db 1003 GlnArgAlaAsnAlaAsnAlaValHisSerMetGluLysValIleGluIleHisSerLys 1022
Qy 2606 CACTGGAGCTGTGTGACAGAGTTCGCGCGTGTGCTGGCGCGTCCCTGCCAGGGGCCAA 2665
Db 1023 TyrTrpHisSerLeuGlnArgTyrAlaSerThrValGlyTyrSerLeuValGluAlaGln 1042

QY 752 CACGGGATGCTCCCTGCTCAGCATGTGCTGTGGAGAGCGGCCGCGCAGCAGC 811
DB 409 ---AlaHisSerProLeuLeuGlnHisMetValLeuLeuGlnProProAlaGlnAla 427
QY 812 ACCCTCAAT-----GCTGTGCCACTCCACGGGAGTCCCTACTAGTACGGGT 859
DB 428 ProLeuValThrGlyLeuGlyAlaLeuProLeuHisAlaGlnSer---LeuValGlyAla 446
QY 860 GAACGTGTGGCCACAGCATGGGACGTAGGACGTCCCGCGGCATCGCCCTGAGC 919
DB 447 AspArgValSerProSer-----IleHisLysLeuArgGlnHisArgProLeuGly 463
QY 920 CGCACTAGTCTCACCGCTCGCGCAGAGTCCCGAGGCGCTCGCAGCAGCTGGTATGCAA 979
DB 464 ArgThrGlnSerAlaProLeuProGlnAsnAlaGlnAlaLeuGlnHisLeuValIleGln 483
QY 980 CAACAGCACCGACAGTCTCTGGAGAAGCAGAG-----CACGACGAGCTACAG 1027
DB 484 GlnGlnHisGlnGlnPheLeuGlnLysHisLysGlnGlnPheGlnGlnGlnLeuGln 503
QY 1028 CTGGGCAAGATCCTCACCAAGACAGGGAGGAGTCCCGAGCAGCCACACCCACCTGAG 1087
DB 504 MetAsnLysIleProLysProSerGluProAlaArgGlnProGluSerHisProGlu 523
QY 1088 GAGCAGAGGAGGAGTACGACGAGCAGGAGGAGTCTGTGGGGAGGAGCCCTG--- 1144
DB 524 GluThrGluGluLeuArgGluHisGln---AlaLeuLeuAspGluProTyrLeuAsp 542
QY 1145 ACCATCCCCGGGAGGCTCCACAGAGATGAGAGCAGACAGAGAGACTGGAGGAGG 1204
DB 543 ArgLeuProGlyGlnLysGluAlaHisAlaGlnAlaGlyValGln---ValLysGlnGlu 561
QY 1205 GACGAGGAAGAGGAGGGAGGAGGAGGAGTTCATCCAGGTTAAGCAGCAGGAGGC 1264
DB 562 ProIleGluSerAspGluGluAlaGlu-----ProProArgGluVal 576
QY 1265 GAGAGTGGTCTCAGAGGGGCGGACTTGGAGGAGCCTGGTGTGATACAAAACTG 1324
DB 577 GluProGlyGlnArgGln---ProSerGluGlnGluLeuLeuPheArgGlnAlaLeu 595
QY 1325 TTCTCAGATGCCACCGCTGCAACCTTTCAGAGGTGTACCAAGCGCCCTCAGCCTGGCC 1384
DB 596 LeuLeuGluGlnGlnArgIleHisGlnLeuArgAsnTyrGlnAlaSerMetGluAlaAla 615
QY 1385 ACTGTGCC-----CACCAAGCCCTGGGCGCTACCCAACTCCTCCCTGCT 1429
DB 616 GlyIleProValSerPheGlyGlyHisArgProLeuSerArgAlaGlnSerSerProAla 635
QY 1430 GCC-----CCTGGGGCATGAAGAACCCCGCAGACCAACCGTCAAGCACCTCTTC 1480
DB 636 SerAlaThrPheProValSerValGlnGluProProThrLysPro-----ArgPhe 652
QY 1481 ACCCAAGTGTGTCTACACACGTTTCATGCTAAAGCACAGTGCATGTGGGGAACACA 1540
DB 653 ThrThrGlyLeuValTyrAspThrLeuMetLeuLysHisGlnCysThrCysGlySerSer 672
QY 1541 CACGTGACCCCTGAGCATGCTGGCCGGATCCAGAGCATGTGGTCCCGGTGAGGAGACA 1600
DB 673 SerSerHisProGluHisAlaGlyArgIleGlnSerIleTyrSerArgLeuGlnGluThr 692
QY 1601 GGCTGCTTAGCAAGTGGCGAGGATCCGAGGTGCGCAAGCCACGCTAGATGATCCAG 1660
DB 693 GlyLeuArgGlyLysCysGluCysIleArgGlyArgLysAlaThrLeuGluLeuGln 712
QY 1661 ACAGTGACCTCAATACACACCTCTCTATGGACAGTCCCTCAACCGGCAGAG 1720
DB 713 ThrValHisSerGluAlaHisThrLeuLeuTyrGlyThrAsnProLeuAsnArgGlnLys 732
QY 1721 CTAGACAGCAAGATGTGCTCGGTCCCATCATGACGACAGAGATGTATGCTGTGCTGCTGT 1780
DB 733 LeuAspSerLysLysLeuLeuLeuGlySerLeuAla---SerValPheValArgLeuProCys 751

QY 1781 GGGGGCATCGGGTGGACAGTGCACACCTGTGGAAATGAGATGCACCTCTCCAGCTGCTGTG 1840
DB 752 GlyGlyValGlyValAlaAspSerAspThrIleTyrAsnGluValHisSerAlaGlyAlaAla 771
QY 1841 CGATGGCATGGGCTGGCTGGAGCTGGCTTCAAGGTGGCTGCAGGAGAGCTCAAG 1900
DB 772 ArgLeuAlaValGlyCysValValGluLeuValPheLysValAlaThrGlyGluLeuLys 791
QY 1901 AATGGATTTCCATATCCCGGCCCGCCAGGACACAGCCGAGGAATCCACAGCATGGGA 1960
DB 792 AsnGlyPheAlaValValArgProProGlyHisHisAlaGluGluSerThrProMetGly 811
QY 1961 TTCTGTCTTCACTCTAGCATCACCGCAAAACTCTACAGCAGAAAGTTGAACGTG 2020
DB 812 PheCysTyrPheAsnSerValAlaValAlaLysLeuLeuGlnGlnArgLeuSerVal 831
QY 2021 GGCAGGTCTCATCGTGGAGTGGACATTCACCATGGCAATGGCACCCAGCGGCTTC 2080
DB 832 SerLysIleLeuIleValAlaAspTyrPaspValHisHisGlyAsnGlyThrGlnGlnAlaPhe 851
QY 2081 TACATGACCCCTCTGTGCTCTACATCTCTGTGCTATGACAAACGGGAATCTCTTT 2140
DB 852 TyrSerAspProSerValLeuTyrMetSerLeuHisArgTyrAspAspGlyAsnPhePhe 871
QY 2141 CCAGGCTCTGGGGCTCTCTGAAGAGTTGGTGGAGGACCGAGCGGTGGGTACAATGTGAAC 2200
DB 872 ProGlySerGlyAlaProAspGluValGlyThrGlyProGlyValGlyPheAsnValAsn 891
QY 2201 GTGCATGGACAGAGGTGGACCCCGCCATTTGGAGACGTGGAGTACCTACAGCCCTC 2260
DB 892 MetAlaPheThrGlyLeuAspProMetGlyAspAlaGluTyrLeuAlaAlaPhe 911
QY 2261 AGGACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2320
DB 912 ArgThrValValMetProIleAlaSerGluPheAlaProAspValValLeuValSerSer 931
QY 2321 GGGTTGATGCTGTGAAGACATCTGTCTCTCTGGGTGGTGGTGGTGGTGGTGGTGGTGG 2380
DB 932 GlyPheAspAlaValGluGlyHisProThrProLeuGlyTyrAsnLeuSerAlaArg 951
QY 2381 TGTTTTGGCCACTTGCACCGCAGTGCATGACCTGGAGGGCGGGGGTGGTGGTGGTGG 2440
DB 952 CysPheGlyTyrLeuThrLysGlnLeuMetGlyLeuAlaGlyGlyArgIleValLeuAla 971
QY 2441 CTGAGGAGGAGCCATGACTTGCACCGCCATCTGTGATGCTCTGAAGCTTGTGTCTGGCT 2500
DB 972 LeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValSerAla 991
QY 2501 CTGCTCAGTGTAAAGCTGCACGCCCTTGGATGAGGAGTCTTTCAGCAAAACCCCAACATC 2560
DB 992 LeuLeuGlyAsnGluLeuAspProLeuProGluLysValLeuGlnGlnArgProAsnAla 1011
QY 2561 AACGAGTGGCCACGCTAGAGAAAGTCTATCGAGATCCAGACAAACACTGGAGCTGTGTG 2620
DB 1012 AsnAlaValArgSerMetGluLysValMetGluIleHisSerLysTyrTyrArgCysLeu 1031
QY 2621 CAGAGTTCGGCGTGGTCTGGCGGCTCCCTGGAGGGCGCCAGGAGGTGAGACCCGAA 2680
DB 1032 GlnArgThrThrSerThrAlaGlyArgSerLeuIleGluAlaGlnThrCysGluAsnGlu 1051
QY 2681 GAAGCCGAAAT--GTGAACGCCATGGCCCTGTGTGGTGGGGCCGGAACAGGCCCAAGCT 2739
DB 1052 GluAlaGluThrValThrAlaMetAlaSerLeuSerValGlyValLysProAlaGlu--- 1070
QY 2740 GCGGCGCCCGGGGAACACAGCCCCAGCGCCGAGGAGGCCCATGGACGAGGAGCCTGCC 2799
DB 1071 -----LysArgProAspGluGluProMetGluGluGluPro 1083
QY 2800 CTG 2802
DB 1084 Leu 1084
RESULT 5

HDA9_HUMAN
ID HDA9_HUMAN STANDARD; PRT; 1011 AA.
AC Q9UKV0; Q94845;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Histone deacetylase 9 (HD9) (HD7B) (HD7).
GN HDAC9 OR HDAC7B OR HDAC7 OR KIAA0744.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 4), AND CHARACTERIZATION.
RX MEDLINE=21438017; PubMed=11535832;
RA Zhou X., Marks P.A., Rifkind R.A., Richon V.M.;
RT "Cloning and characterization of a histone deacetylase, HDAC9.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:10572-10577(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [3]
RP SEQUENCE OF 99-650 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=99435038; PubMed=10523670;
RA Wang A.H., Bertos N.R., Vezmar M., Pelletier N., Crosato M.,
RA Heng H.H., Th'ng J., Han J., Yang X.J.;
RT "HDAC4, a human histone deacetylase related to yeast HDAL, is a
RT transcriptional corepressor.";
RL Mol. Cell. Biol. 19:7816-7827(1999).
CC -1- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
CC THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS. ACT
CC AS A TRANSCRIPTIONAL REPRESSOR.
CC -1- SUBUNIT: Interacts with MEF2.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms; 1 (shown here), 2,
CC 3/HDHP and 4/HDAC9a; are produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY. HD SUBFAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC
CC EMBL; AY032737; AAK66821.1; -
CC EMBL; AY032738; AAK66822.1; -
CC EMBL; AB018287; BAA34464.1; -
CC EMBL; AF124924; AAF04254.1; -
CC MIM; 606543; -
CC InterPro: IPR000286; His_deacetylase.
CC Pfam: PF00850; Hist_deacetyl; 1.
CC PRINTS; PR01270; HDASUPER.
KW Hydrolase; Nuclear protein; Transcription regulation; Repressor;
KW Alternative splicing.
FT VARSPLIC 487 574
FT VARSPLIC 575 590
FT MISSING (IN ISOFORM 2).
FT PFLEPHTHALSVRQA -> VIGKDLAPGVIKVII (IN
FT ISOFORM 3).
FT VARSPLIC 591 1011
FT MISSING (IN ISOFORM 3).
FT GTGLGEGYNINIAWTGGLD -> RFISLEPHEFYLYLSGNCI
FT A (IN ISOFORM 4).
FT

FT VARSPLIC 880 1011 MISSING (IN ISOFORM 4).
FT CONFLICT 99 L -> I (IN REF. 3).
FT CONFLICT 437 T -> P (IN REF. 3).
FT CONFLICT 644 HQCV -> KPNS (IN REF. 3).
SQ SEQUENCE 1011 AA; 111297 MW; 43ED2785E73CD924 CRC64;
Alignment Scores:
Pred. No.: 5,73e-101 Length: 1011
Score: 2579.00 Matches: 540
Percent Similarity: 72.05% Conservative: 107
Best Local Similarity: 60.13% Mismatches: 179
Query Match: 48.24% Indels: 72
DB: 1 Gaps: 19
US-09-502-945-2 (1-2885) x HDA9_HUMAN (1-1011)
QY 2 GAATTCCTCTGTGCAAGTCA-----AAGAGGCCACACACAGCGCGCTCAACCATTC 55
Db 149 GluPheLeuLeuSerLysSerAlaThrLysAspThrProThrAsnGlyLysAsnHisSer 168
QY 56 CTCCACAGACACCCAAATGCTGG-----GGAGCCACCATCTCTCTTTGGACAGAGT 109
Db 169 ValSerArgHisProLysLeuTyrThrAlaAlaHisHisThrSerLeuAspGlnSer 188
QY 110 TCCCTCCCGGAGCGCGCCCTCCCTCTACAACTGCTTTCCTGGCTGG 169
Db 189 SerProLeuSer-----GlyThrSerProSerTyrLysThrLeuProGly 205
QY 170 CCTACGACAGTCGAGACGACTTCCCTCCGCAACACAGCTCTGAACCACTTGA 229
Db 206 AlaGlnAspAlaLysAspPheProLeuArgLysThrAlaSerGluProAsnLeuLys 225
QY 230 GTCCGTTCAAGGCTAAACAGAGGTGGTGACGGAGAGCAGTCCCTCTCGCTCGC 289
Db 226 ValArgSerArgLeuLysGlnLysValAlaGluArgSerProLeuLeuArgArg 245
QY 290 AAGGATGGGACTCTTATAGCACCCTTAAGAGAGAGAGCTGTGAGATCAGAGTGC 349
Db 246 LysAspGlyAsnValValThrSerPheLysLysArgMetPheGluValThr----- 262
QY 350 CTGGGGCGCTGCTGTGTAAACAGCGACCGCGCTCGGGCCCGCAGCTCTCCCAACAGC 409
Db 263 -----GluSerValSerSerSerProGlySerGlyProSerProAsn--- 279
QY 410 TCCACAGACCATCTCTGAGATGCTTACTGGCTAGTCCCAACATCCCACTGAG 469
Db 280 -----AsnGlyProThrGlySerValThrGluAsnGluThrSer 292
QY 470 ATGCTCCCT-----CAGCAGCGAGCGCTCCCTCTG 499
Db 293 ValLeuProThrProHisAlaGluGlnMetValSerGlnGlnArgLeuLeuHis 312
QY 500 GACAGTCCCGCAACAGTTACGCTCTACAGCTCTCTCTCTGCCCCAACATCTCCCTA 559
Db 313 GluAspSerMetAsnLeuLeuSerLeuTyrThrSerProSerLeuProAsnIleThrLeu 332
QY 560 GGCTCTCAGCGCCAGCTGCTACCACTCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 619
Db 333 GlyLeuProAla-----ValProSerGlnLeuAsnAlaSerAsnSerLeuLys 348
QY 620 ACACAGCAGGAGCGCGAGGAGCGCGCTCCAGTCCCTCGCGCAGGCTGGCAGCTGACC 679
Db 349 GluLysGlnLysCysGlu-----ThrGlnThrLeuArgGlnGlyValProLeuPro 365
QY 680 GCAAGTTC-----ATGAGCACATCTCTATTCTCTGGTCTGCTCTGGCGCTG 727
Db 366 GlyGlnTyrGlyGlySerIleProAlaSerSerHisPro-----HisVal 381
QY 728 GACCTGAGGGGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787
Db 382 ThrLeuGluGlyLysProAsnSerSerHisGlnAlaLeuLeuGlnHisLeuLeuLeu 401
QY 788 CTGGACAGCGCGCGCGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 841


```
Db 402 LysGluGlnMetArgGlnGlnLysLeuValAlaGlyValProLeuHisProGln 421
||||| ||||||| |||:||||| ||||||| |||
842 TCCCCACATGACGGGTGAACGTGTGGCCACACAGCATGCGGAGGTAGGCAAGCTCCCG 901
||||| ||| |||||||:||||| |||||||
Db 422 SerProLeuAlaThrLysGluArgIleSerProGlyIleArgGlyThrHisLysLeuPro 441
||||| ||| |||||||:||||| |||||||
Qy 902 CGGCATCGCCCTGACGCGGCACACTCAGTCCCTCAGCGTCCGCGCAGAGTCCCGCCGCTG 961
||||| |||||||:||||| ||||||| |||
Db 442 ArgHisArgProLeuAsnArgThrGlnSerAlaProLeuProGlnSer 459
||||| |||||||:||||| ||||||| |||
Qy 962 CAGCAGCTGGTTCATCAACAACAGCAGCAGAGTTCCTGGAGAACAGCAAG 1015
||||| |||||||:||||| ||||||| |||
Db 460 AlaGlnLeuValIleGlnGlnHisGlnGlnPheLeuGluLysGlnLysGlnTyrGln 479
||||| ||| |||||||:||||| ||||||| |||
Qy 1016 CAGCAGCTACAGCTGGGCAAGATCCTCACAAGACAGGGGAGCTGCCAGGACGCCACC 1075
||||| |||:|||||:||||| |||
Db 480 GlnGlnIleHisMetAsnLysLeuLeuSerLysSerIleGluGlnLeuLysGlnProGly 499
||||| ||| |||||||:||||| |||
Qy 1076 ACCCACCCTGAGACAGAGGAGGAGCTGACGGAGCAGCAGGAGGTCTGCTGGGGGAG 1135
||||| ||||||| ||||||| |||
Db 500 SerHisLeuGluGluAlaGluGlnLeuGln 512
||||| ||| |||||||:||||| |||
Qy 1136 GGAGCGCTGACCATGCCCGGGAGGCTCCACAGAGAGTGAGACAGACAGAACACCTG 1195
||||| ||| |||||||:||||| |||
Db 513 GlnAlaMetGlnGluAspArgAlaProSerSer 531
||||| ||| |||||||:||||| |||
Qy 1196 GAGGAGGAGGACGAGAGATGGGAGGAGGAGGAGGATGTCATCCAGGTAAAGAC 1255
||||| ||| |||||||:||||| |||
Db 532 SerAlaCysValAspAspThrLeuGly 549
||||| ||| |||||||:||||| |||
Qy 1256 GAGGAGGCGCAGAGTGGTCTCAGGAGGGCGCGCTGGAGGAGCGCTGCTCGATAC 1315
||||| ||| |||||||:||||| |||
Db 550 GluProValAspSer 567
||||| ||| |||||||:||||| |||
Qy 1316 AAAAACTGTTCTCAGATGCCCAACCGCTCAACCT 1360
||||| ||| |||||||:||||| |||
Db 568 GlnAlaAlaPheMetGlnGlnProPheLeuGluProThrHisThrArgAlaLeuSerVal 587
||||| ||| |||||||:||||| |||
Qy 1361 TACCAAGCGCCCTCAGCTGGCCACTGTG 1408
||||| ||| |||||||:||||| |||
Db 588 ArgGlnAlaProLeuAlaAlaValGlyMetAspGlyLeuGluLysHisArgLeuValSer 607
||||| ||| |||||||:||||| |||
Qy 1409 CGTACCATCTCCCTCCTCCTCGCCCTGGGGCATGAAGAACCCCCACACCAACCCGTC 1468
||||| ||| |||||||:||||| |||
Db 608 ArgThrHisSerProAlaAlaSerValLeuProHisProAlaMetAspArgProLeu 627
||||| ||| |||||||:||||| |||
Qy 1469 AAGCACCTCTTCAACCAAGTGTGTCTACGACACGTTTCATGCTAAAGCACCAGTGCATG 1528
||||| ||| |||||||:||||| |||
Db 628 GlnProGlySerAlaThrGlyIleAlaTyrAspProLeuMetLeuLysHisGlnCysVal 647
||||| ||| |||||||:||||| |||
Qy 1529 TGGGGGAACACACATGACCTGACATGATGTCGGCGGATCCAGAGATCTGGTCCCGG 1588
||||| ||| |||||||:||||| |||
Db 648 CysGlyAsnSerThrThrHisProGluHisAlaGlyArgIleGlnSerIleTrpSerArg 667
||||| ||| |||||||:||||| |||
Qy 1589 CTCAGGAGACAGCGCTGTAGCAAGTGGAGCGGATCCGAGGTCCGAAAGCAGCGTCA 1648
||||| ||| |||||||:||||| |||
Db 668 LeuGlnGluThrGlyLeuLeuAsnLysCysGlyArgIleGlnArgLysAlaSerLeu 687
||||| ||| |||||||:||||| |||
Qy 1649 GATGAGATCCAGACATGCTCTGAATACCAACACCTGCTCTATGGGACCATGCCCTC 1708
||||| ||| |||||||:||||| |||
Db 688 GluGluIleGlnLeuValHisSerGluHisHisSerLeuLeuTyrGlyThrAsnProLeu 707
||||| ||| |||||||:||||| |||
Qy 1709 AACCGGACAGCTACAGCAGAAGATGCTCGGTCCCATCAGCCAGAGATGATGCT 1768
||||| ||| |||||||:||||| |||
Db 708 AspGlyGlnLysLeuAspProArgIleLeuLeuGlyAspSerGlnLysPhePheSer 727
||||| ||| |||||||:||||| |||
Qy 1769 GTCTGCTGTGGGGCATCGGGGGGAGCAGTGACACCGTGTGGATGAGATGCACTCC 1828
||||| ||| |||||||:||||| |||
Db 728 SerLeuProCysGlyGlyLeuValAspSerAspThrIleTrpAsnGluLeuHisSer 747
||||| ||| |||||||:||||| |||
Qy 1829 TCCAGTGTCTGCGCATGGCAGTGGGCTGCTGCTGGAGCTGGCCTTCAAGTGGCTGCA 1888
||||| ||| |||||||:||||| |||
```

```
Db 748 SerGlyAlaAlaArgMetAlaValGlyCysValIleGluLeuAlaSerLysValAlaSer 767
||||| ||||||| |||:||||| ||||||| |||
Qy 1889 GGAGAGCTAAGATGATTGGCTATCATCCGCGCCACAGCACACACCGCGGAGGAATCC 1948
||||| ||||||| |||:||||| ||||||| |||
Db 768 GlyIleLeuLysAsnGlyPheAlaValValArgProGlyHisHisAlaGluGluSer 787
||||| ||| |||||||:||||| ||||||| |||
Qy 1949 ACAGCCATGGGATCTCTCTTCAACTCTGTAGCCATCACCGCAAACTCTACAGCAG 2008
||||| ||| |||||||:||||| ||||||| |||
Db 788 ThrAlaMetGlyPheCysPheAsnSerValAlaIleThrAlaLysTyrLeuArgAsp 807
||||| ||| |||||||:||||| ||||||| |||
Qy 2009 AAGTTGAACGTGGCAGGCTCTCATCTGAGCTGGGACATTCACCATGCAATGGGACC 2068
||||| ||| |||||||:||||| ||||||| |||
Db 808 GlnLeuAsnIleSerLysIleLeuIleValAspLeuAspValHisHisGlyAsnGlyThr 827
||||| ||| |||||||:||||| ||||||| |||
Qy 2069 CAGCAGCGCTTCTACAATGATGCCCTCTGCTCTACATCTCTCTGCTGCTATGACAAC 2128
||||| ||| |||||||:||||| ||||||| |||
Db 828 GlnGlnAlaPheTyrAlaAspProSerIleLeuTyrIleSerLeuHisArgTyrAspGlu 847
||||| ||| |||||||:||||| ||||||| |||
Qy 2129 GGGAACTCTTTCCAGGCTCTGGGCTCTGAAGAGTGTGGAGGACACCGCGTGGG 2188
||||| ||| |||||||:||||| ||||||| |||
Db 848 GlyAsnPhePheProGlySerGlyAlaProAsnGluValGlyThrGlyLeuGlyGluGly 867
||||| ||| |||||||:||||| ||||||| |||
Qy 2189 TACAATGTGAACGTGGATGGACAGGAGGTGTGACCCCGCCCATTTGGAGAGCTGGAGTAC 2248
||||| ||| |||||||:||||| ||||||| |||
Db 868 TyrAsnIleAsnIleAlaTrpThrGlyGlyLeuAspProMetGlyAspValGluTyr 887
||||| ||| |||||||:||||| ||||||| |||
Qy 2249 CTTACAGCCTTCAGCAGTGTGTCATGCCATTGCCACAGAGTCTCCACTGTGTGGTC 2308
||||| ||| |||||||:||||| ||||||| |||
Db 888 LeuGluAlaPheArgThrIleValLysProValAlaLysGluPheAspProAspMetVal 907
||||| ||| |||||||:||||| ||||||| |||
Qy 2309 CTAGTCTCCGCGCGGTTGTGCTGTGAAGACATCTGCTCTCTGGTGGCTACTCT 2368
||||| ||| |||||||:||||| ||||||| |||
Db 908 LeuValSerAlaGlyPheAspAlaLeuGluGlyHisThrProLeuGlyGlyTyrLys 927
||||| ||| |||||||:||||| ||||||| |||
Qy 2369 GTCACCGCCAGATGTTTGGCCACTTGACAGGAGCTGATGACCTGCGGAGGGGCGGG 2428
||||| ||| |||||||:||||| ||||||| |||
Db 928 ValThrAlaLysCysPheGlyHisLeuThrLysGlnLeuMetThrLeuAlaAspGlyArg 947
||||| ||| |||||||:||||| ||||||| |||
Qy 2429 TGTGCTGCGCCCTGGAGGAGGCGCATGACCTGACCGCATCTGTGATGCTCTGAAGCT 2488
||||| ||| |||||||:||||| ||||||| |||
Db 948 ValValLeuAlaLeuGluGlyHisAspLeuThrAlaIleCysAspAlaSerGluAla 967
||||| ||| |||||||:||||| ||||||| |||
Qy 2489 TGTGCTGCGCTGCTCAGTGTAAAGCTGAAGCTGACGCCCTTGGATGAGGAGCTTTCAGCAA 2548
||||| ||| |||||||:||||| ||||||| |||
Db 968 CysValAsnAlaLeuLeuGlyAsnGluLeuGluProLeuAlaGluAspIleLeuHisGln 987
||||| ||| |||||||:||||| ||||||| |||
Qy 2549 AAGCCCAACATCACCGCAGTGGCCACGCTAGAGAAAGTCATCGAGATCCAGCAGC 2602
||||| ||| |||||||:||||| ||||||| |||
Db 988 SerProAsnMetAsnAlaValIleSerLeuGlnLysIleIleGluIleGlnSer 1005
||||| ||| |||||||:||||| ||||||| |||
RESULT 6
HDA9_MOUSE STANDARD; PRT; 588 AA.
ID HDA9_MOUSE
AC Q99N13; Q9EP72;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone deacetylase 9 (HD9) (HD7B) (Histone deacetylase-related
protein) (MEP2-interacting transcription repressor MTR).
GN HDAC9 OR HDAC7B OR HD9P OR MTR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster / NIH;
RA Zhou X., Richon V.M., Rifkind R.A., Marks P.A.;
RT "Cloning of the mouse HD9P cDNA";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH Swiss; TISSUE=Heart;
```


Db 638 ArgAlaLeuArgIleLeuIleValAspTspAspValHisHisGlyAsnGlyThrGlnHis 657
 QY 2075 GCGTTCACAAAGACCCCTCTGTCTCTACATCTCTGTGATCGCTATGACACGGGGAAC 2134
 Db 658 IlePheGluAspAspProSerValLeuTyrValSerLeuHisArgTyrAspArgGlyThr 677
 QY 2135 TTCCTTCA-----GGCTCTGGGCTCCTGAAGAGTGTGTGGAGGACACAGGCGTGGG 2188
 Db 678 PhePheProMetGlyAspGluGlyAlaSerSerGlnValGlyArgAspAlaGlyIleGly 697
 QY 2189 TACAATGTGAACGTGGCAGACAGAGGTGTGGACCCGCCATCGAGAGCTGGAGTAC 2248
 Db 698 PheThrValAsnValProTspAsnGly-----ProArgMetGlyAspAlaAspTyr 714
 QY 2249 CTTACAGCCTTCAGACAGAGTGGTATGCCCATTCGCCAGAGTCTCTACCTGATGCTGGTC 2308
 Db 715 LeuAlaAlaThrHisArgLeuValLeuProIleAlaTyrGluPheAsnProGluLeuVal 734
 QY 2309 CTAGTCTCGCGCTGTCTGTGAAGGACATCTGTCTCTGTGGGTGCTACTCT 2368
 Db 735 LeuIleSerAlaGlyPheAspAlaAlaGlnGly-----AspProLeuGlyGlyCysGln 752
 QY 2369 GTACCCGCCAGATGTTTGGCCACTTGACAGGCGAGTACCTGGCAGGGGGCGG 2428
 Db 753 ValThrProGluGlyTyrAlaHisLeuThrHisLeuLeuMetGlyLeuAlaGlyGlyArg 772
 QY 2429 GTGGTGTGGCCCTGAGGAGGCGCATCTGACCGCCATCTGTGATGCTCTGAAGCT 2488
 Db 773 IleIleLeuLeuGluGlyTyrAsnLeuAlaSerIleSerGluSerMetAlaAla 792
 QY 2489 TGTGTCTGGCTGTCTAGTAAAGCTGAAAGCTGACCCCTTGGATGAGGACGTCTGCAGCAA 2548
 Db 793 CysThrHisSerLeuLeuGly-----AspProProGlnLeuThrLeuLeuArg 809
 QY 2549 AAGCCCAACATCAACGCGTGGCCAGCTAGAGAAAGTATCGAGATCCAGAGCAACAC 2608
 Db 810 ProProGlnSerGlyAlaLeuValSerIleSerGluValIleGlnValHisArgLysTyr 829
 QY 2609 TGGAGCTGTGTCCAGAGTGTCCCGCTGTGTGGCGGTCTGCGAGGGGCCCAAGCA 2668
 Db 830 Trp-----ArgSerLeuArgLeuSerLysMet 838
 QY 2669 GTGTGAGCCGAAGA 2683
 Db 839 GluAspLysGluGlu 843
 RESULT 9
 HDAL_SCHPO STANDARD; PRT; 687 AA.
 AC P56523;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Histone deacetylase clr3 (cryptic loci regulator 3).
 GN CLR3 OR SPBC800.03.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA MEDLINE=98429513; PubMed=9755190;
 RX Grewal S.I., Bonaduce M.J., Klar A.J.;
 RT "Histone deacetylase homologs regulate epigenetic inheritance of
 RT transcriptional silencing and chromosome segregation in fission
 RT yeast."; 150:563-576(1998).
 RL Genetics [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;

Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
 Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 Gentles S., Godle A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
 Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 Skelton J., Simmonds S., Squares R., Squares S., Stevens K.,
 Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 Borzym K., Langer I., Beck A., Leirach E., Reinhardt R., Pohl T.M.,
 Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 Sipakowski G.V., Ussery D., Barrell B.G., Nurse P.;
 "The genome sequence of Schizosaccharomyces pombe";
 Nature 415:871-880(2002).
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN
 TRANSCRIPTIONAL REGULATION AND CELL CYCLE PROGRESSION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
 FAMILY. HD SUBFAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF064207; AAD05212.1; -;
 DR EMBL; AL391034; CAC01518.1; -;
 DR InterPro; IPR000286; Hist_deacetyl; 1.
 DR Pfam; PF00850; Hist_deacetyl; 1.
 DR PRINTS; PR01270; HDASUPER.
 KW Hydrolase; Nuclear protein.
 SQ SEQUENCE 687 AA; 76792 MW; 6B0E4184A056D899 CRC64;
 Alignment Scores:
 Pred. No.: 6,34e-23 Length: 687
 Score: 705.50 Matches: 154
 Percent Similarity: 56.21% Conservative: 86
 Best Local Similarity: 36.07% Mismatches: 144
 Query Match: 13.20% Indels: 43
 DB: 1 Gaps: 10
 US-09-502-945-2 (1-2885) x HDAL_SCHPO (1-687)
 QY 1385 ACTGTGCCCCACCAAGCCCTGGCGGTACCAATCTCCCTCT-----GCTGCCCTGGG 1438
 Db 22 ThrValThrProTspSerIleLeuLeuThrAsnAsnLysProMetSerGlySerGluAsn 41
 QY 1439 GGCATGAAGAACCCCAAGACCAACCCGCTCAAGCACCTTCCACCACCAAGTGTGGTCTAC 1498
 Db 42 ThrLeuAsnAsnGluSerHisGluMetSerGlnIleLeuLysLysSerGlyLeuCysTyr 61
 QY 1499 GACAGCTTCATGCTAAAGCACACAGTGCATGTGCGGGAACACACACAGTGCACCTGAGCAT 1558
 Db 62 AspProArgMetArgPheHisAlaThrLeu---SerGluValAspAspHisProGluAsp 80
 QY 1559 GCTGCCCGGATCCAGAGCATGTGTCGGCTGCGAGGAGACAGGCGCTCTTAGCAAGTGC 1618
 Db 81 ProArgArgValLeuArgValPheGluAlaIleLysLysAlaGlyTyrValSerAsnVal 100


```
QY 1619 GAG-----CGGATCCGAGGTGCGCAAGCCAGCTAGATGAGATC 1657
Db 101 ProSerProSerAspValPheLeuArgGileProAlaArgGluAlaThrLeuGluGluLeu 120
QY 1658 CAGACAGTGCACCTCT---GAATACCACACCTGCTCTATGGACAGTCCCTCAACCGG 1714
Db 121 LeuGlnValHisSerGlnGluMetTyrAspArgValThrAsnThrGluLysMetSerHis 140
QY 1715 CAGAAGCTAGACAGCAAGATTGCTCGGTCCCATCAGCAGAGATGATGCTGTGCTG 1774
Db 141 GluAspLeuAlaAsnLeuGluLysIle----- 149
QY 1775 CTTGTGGGGGCATCGGGGTGGAGTGCACCGTGTGGAATGAGATGCACTCTCCAGT 1834
Db 150 -----SerAspSerLeuTyrTyrAsnAsnGluSerAlaPhe 161
QY 1835 GCTGTGGCATGCGATGGCTCCCTGCTGGAGCTGGCTTCAAGTGGCTGCAGGAGAG 1894
Db 162 CysAlaArgLeuAlaCysGlySerAlaIleGluThrCysThrAlaValThrGlyGln 181
QY 1895 CTCAAGAATGATTGCCATCATCCGCGCCCGCAGGACACACCGCCGAGGAATCCACGCC 1954
Db 182 ValLysAsnAlaPheAlaValArgProProGlyHisHisAlaGluProHisLysPro 201
QY 1955 ATGGGATTCTGCTTCAACTCTGTAGCATCACCAGCAAACTCCTCAGCAGCAAGTTG 2014
Db 202 GlyGlyPheCysLeuPheAsnValSerValThrAlaArgSerMetLeuGlnArgPhe 221
QY 2015 -----AAGTGGGCAAGTCTCATCTGAGGACCTGGACATTCACCATGGCAATGCACC 2068
Db 222 ProAspLysIleLysArgValLeuIleValAspTrpAspIleHisGlyAsnGlyThr 241
QY 2069 CAGCAGCGCTTCTACAAATGACCCCTGCTGCTACATCTCTGATCGCTATGACAAAC 2128
Db 242 GlnMetAlaPheTyrAspAspProAsnValLeuTyrValSerLeuHisArgTyrGluAsn 261
QY 2129 GGGAACTCTTTCCAGGCTCT-----GGGCTCTCTGAAGAGTGTGGTGGAGCAGCGC 2182
Db 262 GlyArgPheTyrProGlyThrAsnTyrGlyCysAlaGluAsnCysGlyGluGlyProGly 281
QY 2183 GNGGGGTACAATGTACAGTGCATGGACGAGGAGTGTGGACCCGCCCATTCGAGACGTG 2242
Db 282 LeuGlyArgThrValAsnIleProTrpSerCysAla-----GlyMetGlyAspGly 298
QY 2243 GAGTACCTTACAGCTTACAGGACAGTGGTGATGCCATTCGCCACGAGTCTCAGCTGAT 2302
Db 299 AspTyrIleTyrAlaPheGlnArgValValMetProValAlaTyrGluPheAspProAsp 318
QY 2303 GTGGCTCTAGTCTCGCGGGTTTGATGCTGTTGAAGGACATCTCTCTCTCTGGGTGGC 2362
Db 319 LeuValIleValSerCysGlyPheAspAlaAlaLagly-----AspHisIleGlyGln 336
QY 2363 TACTCTGTACCGCCAGATGTTTGGCCACTGACAGCAGCTGATGACCTGGCAGGG 2422
Db 337 PheLeuLeuThrProAlaAlaTyrAlaHisMetThrGlnMetMetGlyLeuAlaAsp 356
QY 2423 GCGCGGGTGTGCTGCGCTGAGGAGGCCATGACTTACCGCCCATCTGTGATGCTCTCT 2482
Db 357 GlyLysValPheIleSerLeuGluGlyGlyTyrAsnLeuAspSerIleSerThrSerAla 376
QY 2483 GAAGCTTGTGCTCGGCTCTGCTAGTGTAAAGCTGCAGCCCTTGGATGAGGAGCTCTTG 2542
Db 377 LeuAlaValAlaGlnSerLeuLeuGlyIleProProGlyArgLeuHisThrThrTyrAla 396
QY 2543 CAGCAAAAGCCCAACATCAACGCACTGCCAGCTGTAGAGAAGTCAAGTCAAGCAGAGC 2602
Db 397 CysProGln-----AlaValAlaThrIleAsnHisValThrLysIleGlnSer 412
QY 2603 AAACACTGGAGCTGTGTCAG 2623
Db 413 GlnTyrTrpArgCysMetArg 419
```

```
RESULT 10
HDAL_YEAST
ID HDAL_YEAST STANDARD; PRT; 706 AA.
AC P53973;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone deacetylase HDAL.
GN HDAL OR YNL021W OR N2819.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Andre B., Iraqui Houssaini I., Urrestarazu L.A., Vissers S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RA MEDLINE-97121415; PubMed-8962081;
RA Rundlett S.E., Carmen A.A., Kobayashi R., Bavykin S., Turner B.M.,
RA Grunstein M.;
RT complexes that regulate silencing and transcription.;
RL Proc. Natl. Acad. Sci. U.S.A. 93:14503-14508(1996).
CC -!- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
CC THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
CC REGULATION AND CELL CYCLE PROGRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY. HD SUBFAMILY 2.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z71297; CAA95883.1;
TRANSFAC; T04600;
SGD; S0004966; HDAL.
InterPro; IPR000286; His_deacetylase.
Pfam; PF00850; Hist_deacetyl; 1.
PRINTS; PR01270; HDASUPER.
KW Hydrolase; Nuclear protein.
SQ SEQUENCE 706 AA; 80069 MW; 4E7069E66D03264D CRC64;
-----
Alignment Scores:
Pred. No.: 1.57e-21 Length: 706
Score: 672.00 Matches: 162
Percent Similarity: 50.40% Conservative: 87
Best Local Similarity: 32.79% Mismatches: 165
Query Match: 12.57% Indels: 80
DB: 1 Gaps: 15
US-09-502-945-2 (1-2885) x HDAL_YEAST (1-706)
QY 1235 GATTGCATCAGTTTAAAGGACGAGGAGGAGTGTGCTGAGGAGGGCGCGACTTG 1294
Db 2 AspSerValMetValLysLysGluVal-----Leu 11
QY 1295 GAGGAGCTGTGTGCTGATACATAAAACTGTCTCAGATGCCCAACCGCTGCACACTTG 1354
Db 12 GluAsnProAspHisAspLeuLysArgLysLeuGluGluAsnLys----- 26
QY 1355 CAGGTGTACCAAGCGCCCTCAGCCTGGCCACTGTGCCCCACCAAGCCCTGGCGGTACC 1414
Db 27 -----GluGluGluAsnSerLeuSerThr-----SerLysSer 38
QY 1415 CAATCTCTCCCTGCTGCCCTGGGGGCGATGAAGAACCCCGGACCAACCCGCTCAAGCAC 1474
```



```
Db 39 LysArgGlnValIleValProValCysMetProLysIleHisTyrSerProLeuLys --- 57
    :
    :
    :
QY 1475 CTCATCACCACAAGTGTGCTACGACACGTTCTATGCTAAAGCACCACGATGTCGGG 1534
    :
    :
    :
Db 58 -----ThrGlyLeuCysTyrAspValArgMetArgTyrHisAlaLysIlePheThr 74
    :
    :
    :
QY 1535 AAC-----ACACAGTCGCACCTCAGCATGCTGGCCGATCCAGAGCATCTGG 1582
    :
    :
    :
Db 75 SerTyrPheGluTyrIleAspProHisProGluAspProArgArgIleTyrArgIleTyr 94
    :
    :
    :
QY 1583 TCCCGCTGTCAGGACAGAGCGCTGCTTACCAAG----- 1615
    :
    :
    :
Db 95 LysIleLeuAlaGluAsnGlyLeuIleAsnAspProThrLeuSerGlyValAspAspLeu 114
    :
    :
    :
QY 1616 -----TGCAGCGGATCCGAGTGCAGACCGTGTGAATGAGATGCATCCACAGTCG 1666
    :
    :
    :
Db 115 GlyAspLeuMetLeuLysIleProValArgAlaIleThrSerGluGluIleLeuVal 134
    :
    :
    :
QY 1667 CACTCTGTAATACAC-----ACCCTGCTCTATGGACACAGTCCCTCAACCGGCAGAGCTA 1723
    :
    :
    :
Db 135 HisThrLysGluHisLeuGluPheIleGluSerThrGluLysMetSerArgGluGluLeu 154
    :
    :
    :
QY 1724 -----GACAGCAAGAAGTGTGCTGGTCCATCAGCAGAGATGTATGCTGTGCT 1777
    :
    :
    :
Db 155 LeuLysGluThrGluLys----- 160
    :
    :
    :
QY 1778 TGTGGGGCATCGGGTGGACATGCACACCGTGTGAATGAGATGCATCCCTCCAGTCG 1837
    :
    :
    :
Db 161 -----GlyAspSerValTyrPheAsnAsnAspSerTyrAlaSer 173
    :
    :
    :
QY 1838 GTCCGATGCGAGTGGCTGCTGGAGCTGGCTTCAAGTGGCTGTCAGGAGAGCTC 1897
    :
    :
    :
Db 174 AlaArgLeuProCysGlyGlyAlaIleGluAlaCysLysAlaValValGluGlyArgVal 193
    :
    :
    :
QY 1898 AGAATGATTTGCCATCATCGGCCCCAGGACACCGCCGAGAGATCCACAGCATG 1957
    :
    :
    :
Db 194 LysAsnSerLeuAlaValValArgProProGlyHisHisAlaGluProGlnAlaGly 213
    :
    :
    :
QY 1958 GGAATCTGCTTCTCAACTCTAGCATCATCAGCCGCAAAA-----CTCCTACAGCAGA 2011
    :
    :
    :
Db 214 GlyPheCysLeuPheSerAsnValAlaValAlaLysAsnIleLeuLysAsnTyrPro 233
    :
    :
    :
QY 2012 TTGAAGCTGGCAAGTCTCATCTGCGGACATTCACCATGGCAATGGCAACCCAG 2071
    :
    :
    :
Db 234 GluSerValArgArgIleMetIleLeuAspTyrPaspIleHisGlyAsnGlyThrGln 253
    :
    :
    :
QY 2072 CAGCGCTTCTACATGACCCCTCTGCTCTCATCTCTGCTGCTGCTATGACACAGGG 2131
    :
    :
    :
Db 254 LysSerPheTyrGlnAspAspGlnValLeuTyrValSerLeuHisArgPheGluMetGly 273
    :
    :
    :
QY 2132 AACTTCTTTCAGGCTCT-----GGGCTCTCTGAAGAGTGTGGGAGGACCGAGCGTG 2185
    :
    :
    :
Db 274 LysTyrTyrProGlyThrIleGlnGlyGlnTyrAspGlnThrGlyGluGlyLysGlyGlu 293
    :
    :
    :
QY 2186 GGCTACAATGTGAACGTGGCATGG-----ACAGGAGGTGTGGACCCGCCCATTTGGAGACGTG 2242
    :
    :
    :
Db 294 GlyPheAsnCysAsnIleThrTrpProValGlyGlyVal-----GlyAspAla 309
    :
    :
    :
QY 2243 GAGTACCTTACACCTTCAGGACAGTGGTGTGATGCCATTCGCCACAGGTCTCACCTGAT 2302
    :
    :
    :
Db 310 GluTyrMetTrpAlaPheGluGlnValValMetProMetGlyArgGluPheLysProAsp 329
    :
    :
    :
QY 2303 GTGCTCTACTCTCCCGCGGTTTGTGCTGTGAAGGACATCTGCTCTCTGCTGGGTGGC 2362
    :
    :
    :
Db 330 LeuValIleIleSerGlyPheAspAlaAlaAspGly-----AspThrIleGlyGln 347
    :
    :
    :
QY 2363 TACTCTCTACCCGCCAGATTTTGGCCACTTGACGAGGAGCTGATGCCCTGGCAGGG 2422
    :
    :
    :
Db 348 CysHisValThrProSerCysTyrGlyHisMetThrHisMetLeuLysSerLeuAlaArg 367
    :
    :
    :
QY 2423 GCGCGGTGTGCTGGCCCTGGAGGAGGAGCCATGACTTGAACGCCCATCTGTGATGCTCT 2482
    :
    :
    :
```

```
Db 368 GlyAsnLeuCysValValLeuGluGlyGlyTyrAsnLeuAspAlaIleAlaArgSerAla 387
QY 2483 GAAGCTTGTCTCTCGGCTCTGCTCAGTGTAAAGCTGCAGCCCTGGATGAGGAGCTTGT 2542
    :
    :
    :
Db 388 LeuSerValAlaLysValLeuIleGlyGluProProAspGluLeuProAspProLeuSer 407
    :
    :
    :
QY 2543 CAGCAAAAGCCCAACATCAACGAGTGGCCAGCTAGAGAAAGTCATCGAGATCCAGAGC 2602
    :
    :
    :
Db 408 AspProLysProGluVal-----IleGluMetIleAspLysValIleArgLeuGlnSer 425
    :
    :
    :
QY 2603 AACACATGGAGCTGTGTCAGAGTTCGCGCTGCTGGTGGC 2644
    :
    :
    :
Db 426 LysTyrTrpAsnCysPheArgArgArgHisAlaAsnSerGly 439
    :
    :
    :
RESULT 11
YLFN_CABEL STANDARD; PRT; 798 AA.
ID YLFN_CABEL
AC Q20295; 1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 89.8 kDa protein F41H10.6 in chromosome IV.
GN F41H10.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peioderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Blanchard M., Bradshaw H.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY. HD SUBFAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U61954; AAB03173.1; -.
CC WormPep; F41H10.6; CE10282.
CC InterPro; IPR000286; His_deacetylase.
CC Pfam; PF00850; Hist_deacetyl; 2.
CC PRINTS; PR01270; HDASUPER.
CC Hypothetical protein; Hydrolase.
CC SEQUENCE 798 AA; 89817 MW; E8C0858001040924 CRC64;
Alignment Scores:
Pred. No.: 1,12e-20 Length: 798
Score: 651.50 Matches: 151
Percent Similarity: 59.42% Conservative: 73
Best Local Similarity: 40.05% Mismatches: 110
Query Match: 12.19% Indels: 43
DB: 1 Gaps: 14
US-09-502-945-2 (1-2885) x YLFN_CABEL (1-798)
QY 1547 CACCTGAGCATGTCGGCGGATCCAGAGCATCTGTCGCCGCTGCAGAGACAGCGCTG 1606
    :
    :
    :
Db 446 HisProGluLysProAlaArgThrArgArgIleLeuLysThrLeuArgGluSerGlyVal 465
    :
    :
    :
QY 1607 CTTACCAAGTGG-----GAGCGGATCCGAGGTGCGAAAGCCAGCTAGATGAGATCCAGACA 1663
    :
    :
    :
Db 466 LeuGluLysCysValAspArgAsnCysGluArgIleAlaThrAsnGluGluIleArgLeu 485
    :
    :
    :
QY 1664 GTGCACTCTGAA-----TACCACACCTGCTTATGGACACGATCCCTCAACCGG 1714
    :
    :
    :
Db 486 ValHisThrLysLysMetLeuGluHis-----LeuArgThrThrGluThrMetLysAsp 503
    :
    :
    :
```



```
QY 1715 CAGAAGCTAGACACAGAAGTGTCTCGTCCCATCAGCCAGAGATGTATGTCTGTG 1774
Db 504 GluGluLeuMetGluGlu-----AlaGluLysGluPhe----- 514
QY 1775 CTTTGTGGGGCATCGGGGTGGACAGTGCACACCGTGTGGATGAGATGCTCCTCAGT 1834
Db 515 -----AsnSerIleTyrLeuThrArgAspThrLeu-----Lys 525
QY 1835 GCTGTGGGCATGGCAGTGGCTGCTCGTGGAGCTGGCC-----TTCAAGTGGCT 1885
Db 526 ValAlaArgLysAlaValAlaValAlaValLeuGlnSerValAspGluIlePheGluLysAsp 545
QY 1886 GCAGGAGAGCTCAAGATGGATTTCATCCATCCGCGCCAGACACACACCGCGAGGAA 1945
Db 546 AlaGlyGln---ArgAsnAlaLeuValIleValArgProGlyHisHisAlaSerAla 564
QY 1946 TCACACCCATGGATTCCTGCTTCACTCTGACCTCAGCCGCAAACTCCTACAG 2005
Db 565 SerLysSerSerGlyPheCysIlePheAsnAsnValAlaValAlaLysTyrAlaGln 584
QY 2006 CAGAAGTTGAACGTGGCGCAAGTCTCATCTGTGCTTACATCTCTGCTATGATGAC 2125
Db 585 ArgArgHisLysAlaLysArgValLeuIleLeuAspTrpAspValHisHisGlyAsnGly 604
QY 2066 ACCAGCAGCGCTTCTACAATGACCCCTCTGTCTTACATCTCTGCTATGATGAC 2125
Db 605 ThrGlnGluIlePheTyrGluAspSerAsnValMetTyrMetSerIleHisArgHisAsp 624
QY 2126 AACGGGAACCTTCTCAGGCTCTGGGCTCTGAA-----GAGTGTGGTGGAGGA 2176
Db 625 LysGlyAsnPheTyrPro---IleGlyGluProLysAspTyrSerAspValGlyGluGly 643
QY 2177 CCAGCGCTGTGACATGTGAAGTGGTGGACAGGAGGTGGACCCCGCATGGA 2236
Db 644 AlaGlyGluGlyMetSerValAsnValProPheSerGly-----ValGlnMetGly 660
QY 2237 GAGTGGAGTACCTTACAGCTTCAGACAGATGTGTGATGCCATGCGCCACGAGTCTCA 2296
Db 661 AspAsnGluTyrGlnMetAlaPheGlnArgValIleMetProIleAlaTyrGlnPheAsn 680
QY 2297 CTTGATGTGCTCTAGTCTCGCGGGTGTGATGCTGTGAAGACATCTGTCTCTCTG 2356
Db 681 ProAspLeuValLeuIleSerAlaGlyPheAspAlaAlaValAsp-----AspProLeu 698
QY 2357 GGTGGCTACTCTCTACCGCGCAGATGTTTGGCCACTGTGACGAGGAGTGTGACCCCTG 2416
Db 699 GlyGluTyrLysValThrProGluThrPheAlaLeuMetThrTyrGlnLeuSerSerLeu 718
QY 2417 GCAGGGGCGGGTGTGTGTCGCGCTGGAGGAGGCATGACTTGACCGGCATCTGTGAT 2476
Db 719 AlaGlyGlyArgIleIleThrValLeuGluGlyGlyTyrAsnLeuThrSerIleSerAsn 738
QY 2477 GCCTCTGAAGCTTGTCTCGGCTCTGCTCAGT---GTAAGTGTGAGCCCTTGGATGAG 2533
Db 739 SerAlaGlnAlaValCysGluValLeuGlnAsnArgSerMetLeuArgArgLeuArgGlu 758
QY 2534 -----GCAGTCTTGCAGCAAAAGCCCAACATCAACGAGTGGCCAGCTCA 2578
Db 759 GluLysGluGlnPheAlaThrLysProGlnLysIleLeuSerSerCysIleLysThrIle 778
QY 2579 GAGAAAGTCATCGAGATCCAGACAAACACTGGAGTGTGTGCGAGAGTTC 2629
Db 779 ArgGluValCysAlaValGlnGlnLysTyrTrpSerIleLeuLysGlyPhe 795
RESULT 12
Y130_ARCFU
ID Y130_ARCFU STANDARD: PRT; 359 AA.
AC O30107;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF0130.
GN AF0130.
```

```
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.;
RL Nature 390:364-370(1997).
CC -!- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF001097; AAB91099.1; -.
DR TIGR; AF0130; -.
DR InterPro; IPR00286; His_deacetylase.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PR01270; HDASUPER.
KW Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 359 AA; 40468 MW; CEC0D294311157AC CRC64;
Alignment Scores:
Pred. No.: 5,85e-14 Length: 359
Score: 490.50 Matches: 129
Percent Similarity: 49.02% Conservative: 71
Best Local Similarity: 31.62% Mismatches: 125
Query Match: 9.18% Indels: 83
DB: 1 Gaps: 18
US-09-502-945-2 (1-2885) x Y130_ARCFU (1-359)
QY 1484 ACAAGTGTGTCTACGACACAGTTCATGCTTAAAGCACCAGTGCATGTGGGGAAACACAC 1543
Db 3 ThrGlyIleValPheHisGluGluTyrLeuLysHisGlu-----GlnSerPro 18
QY 1544 GTGCACCTGTGACATGCTGGCGGATCCAGACATCTGTGTCGGCTCGAGGAGACAGGC 1603
Db 19 ThrHisProGluArgArgGluArgLeuAlaTyrThrMetAspGlnLeuArgGluGly 38
QY 1604 CTGCTTAGCAAGTGGCGGGATC-----CGAGGTGCGCAAGCCACGCTAGATGAG 1654
Db 39 IlePhe---GluSerGluArgIleValLeuLeuGluProPheLysAlaSerLeuGluAsp 57
QY 1655 ATCCAGACAGTGCATCTCTGAATACCAACACCCCTGCTCTATGGGACACGCTCCCTCAACGG 1714
Db 58 ValLeuGluValHisThrGluGluTyrValArgPheLeu----- 70
QY 1715 CAGAACCTAGACACAGAAAGTGTGCTCGTCCATCAGCCAGAAAGATGTATGCTGTGCTG 1774
Db 71 ----GluMetGluSerLysLys----- 76
QY 1775 CCTTGTGGGGCATCGGGGTGGACAGTGCACACC-----GTGTGGAATCAG 1819
Db 1775 CTTTGTGGGGCATCGGGGTGGACAGTGCACACC-----GTGTGGAATCAG 1819
```



```
Db 77 -----GlyGlyIle---IleAspPheAspThrAsnIleProValGlyValPheAspArg 93
QY 1820 ATGCACTCTCCAGTGTGTCGGCATGCGCAGTGGCTGCTGCTGGAGCTGGCTTCAAG 1879
Db 94 -----AlaLeuLeuAlaAlaGlyGlyAlaIleArgAlaAlaGlnAla 107
QY 1880 GTGGTCGAGGAGAGCTCAAGAAATGGATTGCTCATCCGCGCCCGCCAGGACACAGCC 1939
Db 108 ValLeuAsnLysGluCysGluAsnAlaPheAlaMetIleArgProProGlyHisHisAla 127
QY 1940 GAGGAATCCAGCAGCATGGATTCTGCTTCTCACTCTGAGCAGTACCACCGCAAAA--- 1996
Db 128 LysProTyrIleGlyAlaGlyPheCysTyrLeuAsnAsnMetAlaIleMetValLysTrp 147
QY 1997 CTCCTACAGCAAGTTGAACGTGGCAGCTCCTCATCCGTCGGAGTGGAGATCACCAT 2056
Db 148 LeuLeuLysGlnGlyPheGlu-----ArgIleAlaIleLeuAspTrpAspAlaHisHis 165
QY 2057 GCAATGACAGCAGCAGCGCTTACATGACCCCTCTGCTCTACATCTCTCTGAT 2116
Db 166 GlyAspGlyThrGlnGluIlePheTyrAsnAspArgValLeuPheIleSerThrHis 185
QY 2117 CCTATGACACAGGAACCTCTCTCCAGCTCTGGGCTCTCTGAAGAGTGTGGTGAGGA 2176
Db 186 GlnMetPro-----LeuTyrProGlyThrGlyTyrProGluGluCysGlyThrGly 202
QY 2177 CCAGCGTGGGTACAAATGTGAACGTGGCATGCGACAGGAGTGTGGACCCCCCATTTGA 2236
Db 203 LysGlyGluGlyTyrThrValAsnIlePro-----LeuProGlyThrGly 218
QY 2237 GAGCTGAGTACTTACAGCCCTTACAGGACAGTGGTGGATGCCATGCCCACAGGATTTCA 2296
Db 219 AspGluGlyTyrMetMetValIleAspGluIleIleGluProValValAsnGluPheLys 238
QY 2297 CCTGATGTGCTCTAGTCTCCCGCGGTTGATGCTGTTGAAGGACATCTGTCT---CCT 2353
Db 239 ProGlnPheIleAlaIleSerAlaGlyGlnAsp-----AsnHisPheThrAspPro 255
QY 2354 CTGGGTGGTACTCTCTACCCCGCAGATGTTTGGCCACTTACAGCAGCAGTGTGAC 2413
Db 256 IleThrSerLeuAlaLeuProTyrThrAsnLeuGlyIleIleAlaAlaMetAlaGlyPheAspLeu 275
QY 2414 CTGGCA-----GGGCGCGGTGGTGGTGGCTGGCCCTGGAGGAGGCGCATGCTTG 2461
Db 276 MetAlaGluLysHisCysAspGlyArgLeuValAlaValLeuGluGlyTyrSerVal 295
QY 2462 ACC---GCCATCTGATGCCCTCTGAAGCTTGCTGCTCGGCTCTGCTAGTGTAAAGCTG 2518
Db 296 GluGlyAlaLeuProTyrThrAsnLeuGlyIleIleAlaAlaMetAlaGlyPheAspLeu 315
QY 2519 CAGCCCTTG-----GATGAG 2533
Db 316 SerAlaIleArgGluProGluAsnTyrLeuProGluLeuLeuTyrPargLysArgAspSer 335
QY 2534 GCAGTCTTCGACAAAGCCCAACATCAACGCGAGTGGCCAGCTAGAGAAAGTCAATCGAG 2593
Db 336 AlaLeuValLysLeuLysHisAsnIle-----GluAspValLysArg 349
QY 2594 ATCCAGAGCAACACTGGAGCTGT 2617
Db 350 ValHisSerLysTyrTrpLysCys 357
RESULT 13
YB94_METTH
ID YB94_METTH STANDARD; PRT; 331 AA.
AC O27262;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MTH1194.
GN MTH1194.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
```

```
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=980037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
J. Bacteriol. 179:7135-7155(1997).";
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE000887; AAB85683.1; -
CC InterPro; IPR000286; His_deacetylase.
CC Pfam; PF00850; Hist_deacetyl; 1.
CC PRINTS; PR01270; HDASUPER.
CC KW Hypothetical protein; Hydrolase; Complete proteome.
CC SEQUENCE 331 AA; 36722 MW; 489054F32965EDCF CRC64;
Alignment Scores:
Pred. No.: 1,46e-09 Length: 331
Score: 385.00 Matches: 104
Percent Similarity: 47.16% Conservative: 54
Best Local Similarity: 31.04% Mismatches: 131
Query Match: 7.20% Indels: 46
DB: Gaps: 10
US-09-502-945-2 (1-2885) x YB94_METTH (1-331)
QY 1535 AACACACAGTCGACCTCGAGCATGCTGGCGGATCCAGACATCTGTGCTCCGCTGCAG 1594
Db 13 AsnHisGluGlyHisValGluAsnSerGlyArgThrArgAlaIleLeuArgAlaIleGlu 32
QY 1595 GAGACAGCCCTGCTAGCAAGTGCAGCGGATCCGAGGTCGCAAGGTCGCAAGCCACGCTAGATGAG 1654
Db 33 SerSerAspLeuSerProArgPhe-----ValGluProGlyMetAlaGlyIleAspAsp 50
QY 1655 ATCCAGACAGTGCATCTGTAATACACACCTGCTCTATGGGACGAGTCCCTCAACCGG 1714
Db 51 IleLeuMetValHisSerSerThrHisVal-----GluTyrLeuGluValPheAla 67
QY 1715 CAGAAGCTAGACAGCAAGAAGTTCGTCGTCCTCCATCAGCCAGAAAGATGATGCTGTGCTG 1774
Db 61 -----GluTyrLeuGluValPheAla 67
QY 1775 CCTTGTGGGGCATCGGGGTGGACAGTGCACACCGTGTGGAATGAGATGCATCTCCTCCAGT 1834
Db 68 GlyArgGlyGlyTyrLeuAspTyrAspThr---TyrMetThrProGluSerPheSer 86
QY 1835 GCTGTGCCATGGCAGTGGGTGCTGCTGAGCTTCAAGGTGGCTTCAAGGTGGCTGAGGAGAG 1894
Db 87 ValAlaArgLeuSerAlaGlyGly-----AlaMetLeuAlaAlaGluAla 102
QY 1895 CTCAGAATGGA-----TTTGGCATCATCCGCGCCCGCCAGCAGCACCGCCGAGGAATCC 1948
Db 103 LeuArgAspGlyTyrPheSerTyrSerLeuGlyArgProGlyHisHisAlaThrTyrAsp 122
QY 1949 ACAGCCATGGGATTCGTCTTCTTCAACTCTGTAGCATCATCCGCAAAACTCTCTACAGCAG 2008
```



```
Db 123 ArgSerMetGlyPheCysIlePheAsnAsnIleAlaIleAlaIleGluHisAlaArg 142
QY 2009 AAGTTGAACGTGGCAAGCTCTCTGCTGGACTGGGACATTCACCATGGCAATGGCACC 2068
Db 143 AsnLeuGlyValSerArgProLeuValLeuAspPheAspValHisGlyAsnGlyThr 162
QY 2069 CAGCAGGGCTTCAATAGACCCCTCTGTCTCTACATCTCTCTGCTATGACCAAC 2128
Db 163 SerSerIlePheTyArgAspValMetTyIleSerIleHisGln---AspPro 181
QY 2129 GGGAACTTCTCCAGGCTCGGGCTCCTGAAGAGTTGGTGGAGACCGAGGCGGG 2188
Db 182 ArgThrLeuPheProGlyThrGlyPheIleAspGluThrGlySerGlyGluGlyGluGly 201
QY 2189 TACAATGTGAACGTGGCATGACAGAGGTGTGGACCCCTTGGAGAGGTGGAGTAC 2248
Db 202 PheAsnLeuAsnIleProMetProArgly-----SerGlyAsnArgGluTy 217
QY 2249 CTTACAGCTTCAGACAGTGGTGGATGCCCATTTGCCACGAGTTCTCCTACCTGATGTGTC 2308
Db 218 LeuTrpIleLeuGlyMetIleLeuProAlaValLeuGluGlyPheArgProAspMetIle 237
QY 2309 CTAGTCTCCGCGGTTTGAAGTCTTGAAGACATCTGCTCTCTCTGCTGGTACTCT 2368
Db 238 PheValSerAlaGlyPheAspAla-----HisArgArgAspProLeuAlaGluIleMet 255
QY 2369 GTCACCCGCGAGATGTTT-----GGCCACTTGACAGCGCAGCTGATGACCCCTGCA 2419
Db 256 ValAspGluGluPhePheSerTrpIleGlyTrpPheIleHisGln-----Thr 271
QY 2420 GGGGGCGGGTGGTGGCGCTCGAGGAGGCGCATGCTTGCAGCGCATCTGTGATGCC 2479
Db 272 GlyLeuProCysThrAlaValLeuGluGlyGlyTyArgProGluAlaLeuGlyArgSer 291
QY 2480 TCTGAAGCTGTGCTCGGCTCTCAGTCAAGTAAAGTGCAGCCC 2524
Db 292 AsnIleAlaPheMetArgGlyLeuAspGlyGluGluTyArgGluPro 306

RESULT 14
Y245_SYN3
ID Y245_SYN3 STANDARD; PRT; 304 AA.
AC P27202;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein slr0245.
GN SLR0245.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsumoto A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL: D90900; BAA16709.1; --
DR InterPro: IPR000286; His_deacetylase.
DR Pfam: PF00850; Hist_deacetyl; 1.
DR PRINTS: PRO1270; HDASUPER.
KW Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 304 AA; 33649 MW; 69B37CF7EB6360DC CRC64;

Alignment Scores:
Pred. No.: 1.85e-08 Length: 304
Score: 358.50 Matches: 105
Percent Similarity: 47.80% Conservative: 58
Best Local Similarity: 30.79% Mismatches: 138
Query Match: 6.71% Indels: 41
DB: 1 Gaps: 11

US-09-502-945-2 (1-2885) x Y245_SYN3 (1-304)
QY 1487 AGTGTGCTTACGACACGTTTCATGCTAAAGCACACGATGTCGCGGGAACACACAGCTG 1546
Db 3 AlaIleIleTySerAlaGluPheLeuArgHisGlu-----ThrGlyProThr----- 18
QY 1547 CACCTCAGCATGCTGCCGATCCAGAGCATCTGTCCCGGTCGACAGGAGACAGGCGCTG 1606
Db 19 HisProGluCysProAlaArgLeuThrAlaIleAlaThrAlaLeuArgLys----- 35
QY 1607 CTTAGCAAGTGGCAGCGGATCCGAGGTGCGAAAGCCACGCTAGATGATCCAGACAGTG 1666
Db 36 -----MetProGlyAlaAsnTyLeu 42
QY 1667 CACTCTGAATA-CCACACCCTCTCTATGGACACGTCCTCAACCGGAGAGCTAGA 1725
Db 43 HistProGlnLysProSerProValThrTrpAsnLeuAspProTyIleLeuArgCysHis 62
QY 1726 CAGCAAGAAGTGTCTCGTCCCATCAGCCAGAAAGATGTATGCTGCTGCTGTGGGG 1785
Db 63 SerGlnGlu-TyrLeu-----AsnLysLeuAlaLysLeuAlaGluLeuGly 78
QY 1786 CATCGGGGTGGACAGTCACACCGTGTGGATGAGATGACACTCTCCAGTGTCTGTGGCAT 1845
Db 78 yGlySerLeuAspAlaAspThrProValSerPro---GlnSerTyAspValAlaArgLe 97
QY 1846 GGCAGTGGGTGCTGCTGGAGCTGGCTTCAAGTGGTGCAGGAGAGCTCAAGATGG 1905
Db 97 uAlaValArgAlaTrpLeuAspGlyValAspHisVal---LeuAsnGlnArgGluAlaVa 116
QY 1906 ATTGGCCATCATCCGGCCCCCAGCACACCGCCAGGAGATCCACAGCATGGGATCTG 1965
Db 116 lPheValLeuAlaArgProGlyHisHisAlaIleArgAsnThrGlyMetGlyPheCy 136
QY 1966 CTTCTTCAACTCTGTAGCCATCACCCCAAACTCTACAGCAGAAAGTTGAACGTGGGCA 2025
Db 136 sLeuLeuAsnAsnValAlaIleAlaAlaHisTyAlaLeuThrArgProGlyValGluAr 156
QY 2026 GPTCTTCATCTGGTGGACATTCACCATGCAATGCACCCAGCAGCGCTTCTACAA 2085
Db 156 gValAlaIleLeuAspTrpAspValHisHisGlyAsnGlyThrGluAlaLeuValAspHi 176
QY 2086 TGACCCTCTGTCTCTACATCTCTCTGCTATGACAAACGGGAAGTCTTCTTCCAGG 2145
Db 176 sAsnProArgIlePheTyCysSerLeuHisGlnPhePro-----CysTyProGlu 193
QY 2146 CTTCTGGGCTCTGGAAGAGTGTGGAGACCGGCTGGGTGACAAATGTGAAGCTGCG 2205
Db 193 yThrGlyAlaAlaGlyAspArgGlyGlnHisAspAsnVal-----LeuAsnIlePr 210
QY 2206 ATGGACAGAGGTGTGGACCCCTTTCAGAGAGTGGAGTACCTTACAGCTTCAGGAC 2265
Db 210 oLeuLysProGlyGlyAspGlyLysVal-----TyrArgGluAlaPheGluHi 226
QY 2266 AGTGGGTGATGCCCATTTGCCACGAGTCTCACCATGATGTGGTCTCTAGTCTCCGCGGGTT 2325
Db 226 sLysValLeuProPheLeuArgGlnValLysProAspLeuLeuValSerAlaGlyTy 246
```


Tue Mar 25 09:28:36 2003

us-09-502-945-2.rsp

Page 21

Search completed: March 21, 2003, 12:47:14
Job time : 78.3901 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:32:38 ; Search time 51.3908 Seconds
(without alignments)
10793.685 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346

Sequence: 1 ggaattcctcttcgaagt.....aaacaaaagttaaaattt 2885

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09502945/runat_14032003_101059_19127/app_query.fasta_1.10979
-DB=PIR_73 -OPMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945 @CGN_1.1.263 @runat_14032003_101059_19127 -NCPU=3
-NO_XLPXY -NO_MMAY -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -FPGAPEXT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FPGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1831.5	34.3	878	2 T17245	hypothetical prote
2	1034	19.3	796	2 T32425	hypothetical prote
3	781.5	14.6	1063	2 T46284	hypothetical prote
4	771	14.4	1095	2 T13964	probable histone d
5	705.5	13.2	687	2 T43797	probable histone d
6	672	12.6	706	2 S62933	hypothetical prote
7	490.5	9.2	359	1 B62666	acetylpolymine am
8	488	9.1	782	2 T22134	hypothetical prote
9	487.5	9.1	310	1 A70481	acetoin utilizatio
10	438.5	8.2	337	2 AB3440	acetylsermidine d
11	432.5	8.1	359	2 T05998	hypothetical prote
12	420	7.9	311	2 D97450	histone deacetylas
13	420	7.9	311	2 AF2668	deacetylase (impor
14	406.5	7.6	517	2 T27101	hypothetical prote

15	389	7.3	380	2	D83174	probable acetylpol
16	386.5	7.2	498	2	F88359	protein Y51H1A.5 l
17	385	7.2	331	1	C69026	acetylpolymine am
18	374.5	7.0	296	2	G87505	histone deacetylas
19	371.5	6.9	335	1	H71071	hypothetical prote
20	358.5	6.7	304	1	S74557	acetylpolymine am
21	357.5	6.7	327	2	G90139	deacetylase, proba
22	352.5	6.6	338	2	H84173	acetoin utilizatio
23	352	6.6	334	2	B75095	probable histone d
24	348	6.5	343	1	G64366	acetylpolymine am
25	339.5	6.4	306	2	AE1911	hypothetical prote
26	338	6.3	425	2	G86217	protein T27G7.14 l
27	335.5	6.3	389	2	C89958	acetoin / acetoin
28	331.5	6.2	387	1	S39643	probable acetylpol
29	319	6.0	320	2	G27762	hypothetical prote
30	312	5.8	461	2	T20163	acetylpolymine am
31	309	5.8	351	2	C90142	hypothetical prote
32	308	5.8	465	2	T23963	acetoal dehydrogen
33	305.5	5.7	389	2	E84054	transcription regu
34	303.5	5.7	433	1	S22284	probable protein homol
35	303	5.7	480	1	S60381	probable deacetyl
36	301	5.6	452	2	S64211	histone deacetyl
37	297.5	5.6	434	2	T11643	histone deacetyl
38	297	5.6	369	2	F81178	histone deacetyl
39	292.5	5.5	405	2	T40300	hypothetical prote
40	292.5	5.5	839	2	F75518	hypothetical prote
41	291	5.4	385	2	A81926	probable histone d
42	288	5.4	513	2	T01413	BHLF1 protein - hu
43	279.5	5.2	660	1	Q0BE3	probable acetylpol
44	273.5	5.1	344	2	B83605	hypothetical prote
45	268	5.0	507	2	T19067	

ALIGNMENTS

RESULT 1

T17245

hypothetical protein DKF2p586J0917.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, September 1999

A:Reference number: Z18722

A:Accession: T17245

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-878 <KOE>

A:Cross-references: EMBL:AL117455

A:Experimental source: adult uterus; clone DKF2p586J0917

C:Genetics:

A:Note: DKF2p586J0917.1

Alignment Scores:

Pred. No.: 1-6e-88 Length: 878
Score: 1831.50 Matches: 424
Percent Similarity: 56.03% Conservative: 124
Best Local Similarity: 43.35% Mismatches: 249
Query Match: 34.26% Indels: 182
DB: 2 Gaps: 22

US-09-502-945-2 (1-2885) x T17245 (1-878)

QY 11 TTGTGGAAGTCAAGAGGCCACACAGCGGCCTCAACCATTCCTCCACAGACACCC 70

Db 27 LeuGluArgThrValHisProAsnSerProGlyIleProTyrArgThrLeuGluProLeu 46

QY 71 AAATGCTGGGAGGCCACCATCTCTTTGGACACAGAGTTCCCTCCACAGACGCGCC 130

Db 47 GluThrGluGlyAlaThrArgSerMetLeuSerPheLeuProValProSerLeu 66

QY 131 CCTGGGACGCCCTCCCTCCCTACAACTGCTTTTGCCTGGCCCTACGAGTCGAGACGAC 190

Db 67 ProSerAspProProGluHis----- 73
Qy 191 TCCCTCCCTCCCAACAGCCTCTGAACCCAACTTGAAGTGCCTCAAGCTAAACAG 250
Db 74 PheProLeuArgLysThrValSerGluProAsnLeuLysLeuArgTyrLysProLys--- 92
Qy 251 AAGTGGCTGAGCGAGAGAGCAGTCCCTCCTCGTGCCTCCCAAGATGGGACTGTATTAGC 310
Db 93 LysSerLeuGluArgArgLysAsnProLeuLeuArgLysGluSerAlaPro-----Pro 110
Qy 311 ACCTTTAAGAGAGAGCTGTGTGATACAGATACAGGTGCCGGGCTGGGGCTCCTCGTGTGT 370
Db 111 SerLeuArgArgProAlaGluThrLeuGlyAspSerSerProSerSerSer----- 128
Qy 371 AACAGCGCCCGGCTCGGGCCCGAGCTCCACACACTCCACACACTCCACAGCACCATCGCTGAG 430
Db 129 ---SerThrProAlaSerGlyCysSerProAsnAspSerGluHis----- 143
Qy 431 AATGGCTTTACTGGCTCAGTCCCAACATCCCACTAGATGCTCCTCCCTCAGCAGCGAGCC 490
Db 144 -----GlyProAsnProIleLeuGlySerGluAlaLeuLeuGlyGlnArgLeu 159
Qy 491 CTCCTCTGAGACAGTCCCGCCCAACAGTTACAGCTCTACAGTCTCTCTCTGCGCCCAAC 550
Db 160 ArgLeuGlnGluThrSerValAlaProPheAlaLeuProThrValSerLeuLeuProAla 179
Qy 551 ATCTCCCTAGGCTGAGCGCCAGGTGCTACTGTACCAACTCACACTCCTCCTCCTCCCG 610
Db 180 IleThrLeuGlyLeuProAla----- 186
Qy 611 AAGTGTCTGACACAGCAGGAGCGCCGAGGAGCGCCCTCCAGTCCCTGCGGAGGTGGC 670
Db 187 -----ProAlaArgAlaAspSerArgArgThrHisProThrLeuGlyProArgGly 204
Qy 671 ACGCTGACCGGCAAGTTATGACACATCTCTATTCCTGGCTGCTGCTGGCGGTGGCA 730
Db 205 ProIleLeuGlySerProHisThrProLeuPheLeuProHis-----GlyLeuGlu 221
Qy 731 CTGGAGGGGAGCGGAGCGCCCGGCGATGCTCCCTGCTCGACATGCTGTGTGTGTG 790
Db 222 ProGluAlaGlyGlyThr-----LeuProSerArgLeuGlnProIleLeuLeuLeu 238
Qy 791 GAGCAGCGCCGCGAGCAGACCCCTCATTTGCTGTGCCACTCCAGGGGAGTCCCACTA 850
Db 239 AspProSerGlySerHisAlaProLeuLeuThrValProGlyLeuGlyProLeuPhe 258
Qy 851 GTACGGGTAAACGTGTGCCACACAGCATCGGAGCGGTAGCAAGTCCCGCGGATCGG 910
Db 259 HisPheAlaGlnSerLeuMetThrThrGluArgLeuSerGlySer---GlyLeuHisTrp 277
Qy 911 CCCCTGAGCGGCACTGCTCCTCAGCGTCCGCGCAGAGTCCCGGAGTCCCGGCGC----- 958
Db 278 ProLeuSerArgThrArgSerGluProLeuProProSerAlaThrAlaProProPro 297
Qy 959 -----CTGAGCAGCTGGTTCATGCCAACACAGCAGCAGCAGTTCCTCG 1000
Db 298 GlyProMetGlnProArgLeuGluGlnLeu-----LysThrHisValGlnValIle 314
Qy 1001 GAGAAGCAGAGCAGCAGCTACAGTGGGCAAGATCCTCACCAGAGACAGGGGAGTGG 1060
Db 315 LysArgSer-----AlaLysProSerGluLys 323
Qy 1061 CCCAGCAGCCACCACCCCTGAGGAGCAGAGAGGAGGAGTGCAGGAGCAGCAGAG 1120
Db 324 ProArg----- 325
Qy 1121 GTCTTGTGGGGAGGAGCGCTGACCATGCTCCCGGAGGAGGCTCCACAGAGTGGAGC 1180
Db 325 ----- 325
Qy 1181 ACACAGGAACCTGGAGGAGGAGGAGCAGGAGAGAGGATGGGAGGAGGAGGATTGC 1240
Db 326 LeuArgGlnIleProSerAlaGluAspLeuGluThrAspGly----- 339

Qy 1241 ATCCAGGTTAAGCAGCAGGAGGCGGAGAGTGGT-----CTGAGGAGGCGCGAGCTTG 1294
Db 340 -----GlyGlyProGlyGlnValValAspAspGlyLeuGluHis 352
Qy 1295 GAGAGCCTGCTGCTGATACAAAAAACTTCTCAGATGCCCAACCGCTGCAACCTTTG 1354
Db 353 ArgGluLeuGlyHisGlyGln-----ProGluAlaArgGlyProAlaProLeu 368
Qy 1355 CAGGTGTACCAAGCGCC----- 1372
Db 369 Gln-----GlnHisProGlnValLeuLeuTrpGluGlnArgLeuAlaGlyArgLeu 386
Qy 1373 -----CTAGCCTGGCCACTGTGCCCCACACAGCC 1402
Db 387 ProArgGlySerThrGlyAspThrValLeuLeuProLeuAlaGlnGlyHisArgPro 406
Qy 1403 CTGGGCGGTACCCAACTCTCCCTGCTGCCCTGGGGCATGAAGACCC----- 1453
Db 407 LeuSerArgAlaGlnSerSerProAlaAlaProAlaSerLeuSerAlaProGluProAla 426
Qy 1454 -----CCAGACCAACCCGCTCAAGCACCTC---TTCACCACA 1486
Db 427 SerGlnAlaArgValLeuSerSerSerGluThrProAlaArgThrLeuProPheThr 446
Qy 1487 AGTGTGTCTACGACAGCTTCTAAAGCACCACTGCATGTCGGGAAACACACACAGTG 1546
Db 447 GlyLeuIleTyrAspSerValMetLeuLysHisGlnCysSerCysGlyAspAsnSerArg 466
Qy 1547 CACCTGAGCATGTGCGCGGATCCAGAGCATCTGGTCCCGGCTGCAGGACAGGCGCTG 1606
Db 467 HisProGluHisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnGluArgGlyLeu 486
Qy 1607 CTAGCAGTCCGAGCGGATCCGAGGTCGCAAGCGCTAGATGATGATGATCCAGACAGTG 1666
Db 487 ArgSerGlnCysGluCysLeuArgGlyArgLysAlaSerLeuGluGluLeuGlnSerVal 506
Qy 1667 CACTCTGAATACCACACACCTCTCTATGGGACCACTCCCTCAACCGGACAGACTAGAC 1726
Db 507 HisSerGluArgHisValLeuLeuTyrGlyThrAsnProLeuSerArgLeuLysLeuAsp 526
Qy 1727 AGCAAGAGTTGCTGCGTCCCATCAGCCAGAGATGATGCTGCTGCTGCTGCTGGGGG 1786
Db 527 AsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPheValMetLeuProCysGlyGly 546
Qy 1787 ATCGGGGTGGACAGTGCACCGTGTGGATGAGATGATGATGATGATGATGATGATGATG 1846
Db 547 ValGlyValAlaSerThrAspThrIleTrpAsnGluLeuHisSerSerAsnAlaAlaArgTrp 566
Qy 1847 GCAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1906
Db 567 AlaAlaGlySerValThrAspLeuAlaPheLysValAlaSerArgGluLeuLysAsnGly 586
Qy 1907 TTGTCATCATCCGCGCCCGGACACCGCCGAGGAGTCCACAGCATGGATTTCTGTC 1966
Db 587 PheAlaValValArgProGlyHisHisAlaAspHisSerThrAlaMetGlyPheCys 606
Qy 1967 TTCCTCACTCTGTAGCCATCACCACCAAACTCCCTACAGCAGAGACTTCAACGTGGGCAAG 2026
Db 607 PhePheAsnSerValAlaIleAlaCysArgGlnLeuGlnGlnGlnSerLysAlaSerLys 626
Qy 2027 GTCTTCATCTGTGGAGTGGGACATTCACCATGGCATGGCAGCCAGCGCTTCTACAA 2086
Db 627 IleLeuIleValAspTrpAspValHisHisGlyAsnGlyThrGlnGlnThrPheTrpGln 646
Qy 2087 GACCCTCTGTGCTTACATCTCTCTGATCGCTATGACAGGGAAGTCTTCTTCCAGGC 2146
Db 647 AspProSerValLeuTyrIleSerLeuHisArgHisAspAspGlyAsnPhePheProGly 666
Qy 2147 TCTGGGCTCTGAGAGAGTGGTGGAGGAGGAGGCTGGGTCACTCAAGTGAACCTGGCA 2206
Db 667 SerGlyAlaValAspGluValGlyAlaGlySerGlyGlyGlyPheAsnValAsnValAla 686

149	SerSerProHisPheGluPro-----TyrArgLeuProThrSerLeuAlaAsnAlaHis	166
QY	176 GAC-----ACTGAGACGACTTCCCTCCGGAACACACGCTCTGAAACCCACTTTGAA	229
DB	167 AsnLeuGlnGlnAlaSerGluPheGlnLeuArgLysValAsnSerGluProAsnLeuLys	186
QY	230 GTGCGTTCAAGGCTAAACAGAAAGTGGCTCAGCGAAGACAGTCCCTCTCGCTCGCTCGC	289
DB	187 Met-----ArgIleArgAlaLysLeuLeuSerLysGlySerSerProValGlnHisVal	204
QY	290 AAGGATGGGACTGTTATTAGCACCTTTAAGACAGAGAGCTCTTGAGATCACAGGTGCGGG	349
DB	205 GlnGlnAsnAsnSerGlnPheAsnPhe-----	213
QY	350 CTGGGGCGTCCGTCGGTGTAAACAGCGCACCCGGCTCCGGCCCCAGCTCTCCCAACAGC	409
DB	214 -----ThrHisProGlnLeuLysArgSerAspSerGluThr	225
QY	410 TCCACAGCACACCATCGTGTGAATGGCTTTACTGGCTCAGTCCCAACACTCCCACTGAG	469
DB	226 SerGlnAsnValProLeuAspPheMetGlnSerSerSerGlnThrAsnLeuProHisLeu	245
QY	470 ATGCTCCCTCAGCACCGACCGCTCCCTCTCGGACAGCTCCCGCCCAACAGTTTCAGCCTCTAC	529
DB	246 MetLeuPro-----	248
QY	530 ACGTCTCTTCTGCCCCAACATCTCCCTAGGG-----	562
DB	249 --SerProSerLeuProAsnLeuAlaAlaGlyAlaPheHisGlyLeuAsnLeuPro	267
QY	563 -----CTGAGGCGCAGGTCTACTGTCAACACTCACACCTCACTGCCTCCCGG	610
DB	268 ValGlyGlnAspLeuAsnAlaPheMetAlaValAlaAsn-----LeuSerPro	283
QY	611 AAGCTGTGCGACAGCAGGAGGCGGAGAGCGACGCGCCCTCCAGCTCTCGGCGGAGGGTGGC	670
DB	284 PheLeuSerLeuProSerLeuLeuAsnLysLys-----LeuGluLeuGlyGly	299
QY	671 ACGCTGACCGCAAGTTTCATGAGCATCTCTATTCTCGGCTGCTGGCGCGTGGCA	730
DB	299 -----	299
QY	731 CTGAGGCGGACGGGACCCCGGCGCATGCTCCTCGTGTGAGCATGTGCTGTTGTG	790
DB	300 LeuThrAspGluGlyAspArgAsnGly-----	308
QY	791 GAGCAGCGCGGACGAGCAGCACCCCTCATTTGCTGCCACTCCACGGGCGAGTCCCACTA	850
DB	309 -----LeuIleGlySerSer-----	313
QY	851 GTGACGGGTGAACGTGTGGCCACCAAGCATGCGGACGGTATGGCAAGCTCCCGCGGCATCGG	910
DB	313 -----	313
QY	911 CCCGTGAGCGGCATCTCAGTCTCACCCTGCGCAGAGTCCCGAGGCCCTGTGAGCAGCTG	970
DB	314 -----SerThrSerSerLeuAlaSerAsn	321
QY	971 GTCATGCAACACAGCACCAAG-----CAGTTCTTGAGAACCAAGCAGCAGCAGCTACAG	1021
DB	322 ValSerMetGlySerHisGlnTyrGlnSerLeuLeuLysGlnGlnIleArgAspLeuVal	341
QY	1028 CTGGGCAAGATCTTCACCAAGACAGGGGAGCTGCCAGGCGAGCCCAACCCACCCCTGAG	1081
DB	342 Leu-----	342
QY	1088 GAGACAGAGGAGGTGACGGAGCAGCAGGAGGTCTTGCTGGGGAGGAGGCCCTGACC	1147
DB	342 -----	342
QY	1148 ATGCCCCGGGAGGCTCCACAGAGAGTGAGAGCACACAGGAAGACCTGGAGGAGAGGAC	1207
DB	343 -----ArgArgLysSerLeuValArgGluAsp	351

Db	683	g-----ValMetLysValGluAspArgGluGlyProSerSerLysLeuAsp	700
Qy	2664	----AAGCAGGTGAGACCGAAGAACCGCAATGT-----GAACG	2698
Db	700	lYsLysAlaProGlnProAlaYsProArgLeuAlaGluArgMetThrThrArgGlu	720
Qy	2699	CCATGGCCTTCTGTGTGGTGGGGCCGA---ACAGGCCCAAGCTCGGC---	2746
Db	720	sLysValLeuGluAlaGlyMetGlyLysValThrSerAlaSerPheGlyGluGluSer	740
Qy	2747	CCCGGGA-----ACACAGCCCCAGCGCCGAG	2773
Db	740	rProGlyGlnThrAsnSerGluThrAlaValAlaLeuThrGlnAspGlnProSerG	760
Qy	2774	AGGAGCCCATGGAGCAGGAGCCTGCCCTGTGACG	2807
Db	760	uAlaAlaThrGlyGlyAlaThrLeuAlaGlnThr	771
RESULT 4			
TI3964			
Probable histone deacetylase (EC 3.5.1.-) HDA2 - mouse			
C:Species: Mus musculus (house mouse)			
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 12-Nov-1999			
C:Accession: Tl3964			
R:Verdel, A.; Khochbin, S.			
J. Biol. Chem. 274, 2440-2445, 1999			
A:Title: Identification of a new family of higher eukaryotic histone deacetylases			
A:Reference number: Z17841; MUID:99107904; PMID:9891014			
A:Accession: Tl3964			
A>Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: mRNA			
A:Residues: 1-1095 <VER>			
A:Cross-references: EMBL:AF006603; NID:g4165860; PID:g4165861; PIDN:AAD0983			
A:Experimental source: strain C57BL/6J, total fetus			
C:Keywords: hydrolase			
Alignment Scores:			
Pred. No.:	5,21e-33	Length:	1095
Score:	771.00	Matches:	182
Percent Similarity:	51.29%	Conservative:	77
Best Local Similarity:	36.04%	Mismatches:	144
Query Match:	14.42%	Indels:	102
DB:	2	Gaps:	15
US-09-502-945-2 (1-2885) x TI3964 (1-1095)			
Qy	1187	GAAGACTGGAGCAGGACGACGAGGAAGAGGATGGGAGGAGGAGGAGGATTCATCCAC	1246
Db	381	GluValLeuGluArgSerValGluThrGluGluAspGluValGluGluAlaVal---	399
Qy	1247	GTTAAGCAGCAGGAGCGCAGAGTGGTGTGAGAGGGGGCCGACTTGGAGGAGCCTGGT	1306
Db	400	LeuGluGluGluGluGluGluGlyGlyTrpGluAla-----	411
Qy	1307	GCTGGATACAAAACACTGTTTCAGATGCCAACCGCTGCAACCTTTGCAGGTGTACCAA	1366
Db	412	-----ThrAlaLeuProMetAspThrTrpProLeu-----	421
Qy	1367	GCGCCCTCAGCCTGGCCACTGTGCCCCACCAAGCCCTGGGCCGTACCCCAATCTCCCT	1426
Db	421	-----	421
Qy	1427	GCTGCCCTGGGGCATGAAGAACCCGCCAGACCAACCCGTCACCAACCTCTTCACCACA	1486
Db	422	-----LeuGlnAsn-----ArgThr	426
Qy	1487	AGTGTGTCTACGACAGCTTCATGCTAAAGCACCAGTCATGTGCGGGAACACACACGTG	1546
Db	427	GlyLeuValTyraAspGluLysMetMetSerHis---CysAsnLeuTrpAspAsnHis---	444
Qy	1547	CACCTTGAGCATCTCGCGGATCCAGAGCATCTGTGCCCGCTGCAGGAGACAGCGCTG	1606
Db	445	HispGluThrProGlnArgIleLeuArgIleMetCysHisLeuGluGluValGlyLeu	464


```
Db 182 VallyAsnAlaPheAlaValValArgProProGlyHisHisAlaGluProHisLysPro 201
:::||||| |||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 1955 ATGGGATTCCTGCTTCAACTCTGTAGCCATCACCGCAAACTCTACAGCAGAGTTG 2014
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 202 GlyGlyPheCysLeuPheAsnValSerValAlaArgSerMetLeuGlnArgPhe 221
:::||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2015 -----AAGTGGGCAAGTCTTCATCGTGGACTGGGACATTCACCATGGCAATGCACC 2068
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222 ProAspLysIleLysArgValLeuIleValAspTyrAspIleHisHisGlyAsnGlyThr 241
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2069 CAGCAGCGCTTACAAATGACCCCTCTGCTCTACATCTCTGATCGCTATGACAAC 2128
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 GlnMetAlaPheTyrAspAspProAsnValLeuTyrValSerLeuHisArgTyrGluAsn 261
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2129 GGAACCTCTCTTCCAGGCTCT-----GGGCTCTCGAAGAGTTCGTGGAGACAGGC 2182
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 GlyArgPheTyrProGlyThrAsnTyrGlyCysAlaGluAsnCysGlyGluGlyProGly 281
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2183 GTGGGTACAATGTGAACGTGGCATGGACAGGAGGTGTGGACCCCTTGGAGACGTG 2242
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 282 LeuGlyArgThrValAsnIleProTyrSerCysAla-----GlyMetGlyAspGly 298
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2243 GAGTACCTTACAGCTTCAGGACAGTGGTGATGCCCATTTGCCACGAGTTCTCACTGAT 2302
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 299 AspTyrIleTyrAlaPheGlnArgValMetProValAlaTyrGluPheAspProAsp 318
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2303 GTGGTCTAGTCTCCCGCGGTTTGATGCTGTGTGAAGGACATCTGTCTCTCGGTGGC 2362
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 319 LeuValIleValSerCysGlyPheAspAlaAlaGly-----AspHisIleGlyGln 336
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2363 TACTCTGTCCAGCCGACAGATGTTTGGCCACTTGACAGCAGCTGATGACCTGGCAGGG 2422
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 337 PheLeuLeuThrProAlaAlaTyrAlaHisMetThrGlnMetLeuMetGlyLeuAlaAsp 356
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2423 GCGCGGTGGTCTGCGCTGGAGGAGGCCATGATGTGACCGCCATCTGTGATGCTCT 2482
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 GlyLysValPheIleSerLeuGluGlyGlyTyrAsnLeuAspSerIleSerThrSerAla 376
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2483 GAAGCTGTGTCTCGCTCTGCTCAGTGTAAGCTGCACGCCCTTGGATGAGCAGCTCTTG 2542
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 377 LeuAlaValAlaGlnSerLeuLeuGlyIleProGlyArgLeuHisThrThrTyrAla 396
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2543 CAGCAAAAGCCCAACATCAACAGCTGGCCAGCTAGAGAAAGTATCGAGATCCAGAGC 2602
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 397 CysProGln-----AlaValAlaThrIleAsnHisValThrLysIleGlnSer 412
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2603 AAACACTGGAGCTGTGTGCAG 2623
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 413 GlnTyrTyrArgCysMetArg 419
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 6
S62933
hypothetical protein YNL021w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N2819
C:Species: Saccharomyces cerevisiae
C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 16-Dec-1998
C:Accession: S62933
R:Andre, B.; Iraqui Houssaini, I.; Urrestarazu, L.A.; Vissers, S.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62920
A:Accession: S62933
A:Molecule type: DNA
A:Residues: 1-706 <AND>
A:Cross-references: EMBL:Z71297; NID:gl301850; PID:e239664; PTD:gl301851; MIPS:YNL021w
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:HDAL
A:Cross-references: SGD:S0004966; MIPS:YNL021w
A:Map position: 141
C:Superfamily: RPD3/acuC homology
F:51-388/Domain: RPD3/acuC homology <RAHI>
```

```
Alignment Scores:
Pred. No.: 7,73e-28 Length: 706
Score: 672.00 Matches: 162
Percent Similarity: 50.40% Conservative: 87
Best Local Similarity: 32.79% Mismatches: 165
Query Match: 12.57% Indels: 80
DB: 2 Gaps: 15

US-09-502-945-2 (1-2885) x S62933 (1-706)
QY 1235 GATTGCATCCAGTTAAGGACGAGGAGGCGAGAGTGTGCTGAGGAGGGGCCGACTTG 1294
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 AspSerValMetValLysLysGluVal-----Leu 11
||| ||| ||||| ||||| ||||| ||||| |||||
QY 1295 GAGGAGCGTGGTGGATACAAAACCTGTTCTCAGATGCCAACCCGCTGCAACCTTTG 1354
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12 GluAsnProAspHisAspLeuLysArgLysLeuGluGluAsnLys----- 26
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1355 CAGGTGTACCAAGCGCCCTCAGCCTGGCCACTGTGCCCCACCAAGCCCTGGCCGCTACC 1414
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27 -----GluGluGluAsnSerLeuSerThrThr-----SerLysSer 38
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1415 CAATCCTCCCTGCTGCCCTGGGGGCATGAAGAACCCGCCAGACCAACCCGCTCAAGCAC 1474
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39 LysArgGlnValIleValProValCysMetProLysIleHisTyrSerProLeuLys--- 57
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1475 CTCCTCACCAAGTGTGTCTACGACACGTTCTATGCTTAAAGCACCACTGTCATGTCGG 1534
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 58 -----ThrGlyLeuCysTyrAspValArgMetArgTyrHisAlaLysIlePheThr 74
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1535 AAC-----ACACAGCTGCACCTGAGCATGTGCGCGGATCCAGCATCTGG 1582
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 SerTyrPheGluTyrIleAspProHisProGluAspProArgArgIleTyrArgIleTyr 94
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1583 TCCCGCTGCAGGACAGAGCTGCTTAGCAAG----- 1615
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95 LysIleLeuAlaGluAsnGlyLeuIleAsnAspProThrLeuSerGlyValAspAspLeu 114
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1616 -----TGGAGCGGATCCGAGGTGCGAAAGCCAGCTAGATGATGACACAGCTG 1666
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 115 GlyAspLeuMetLeuLysIleProValArgAlaAlaThrSerGluGluLeuGluVal 134
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1667 CACTCTGAATACCAAC---ACCCTGCTCTATGGGACCACTCCCTCAACCCGCCAAGCTA 1723
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 HisThrLysGluHisLeuGluPheIleGluSerThrGluLysMetSerArgGluLeu 154
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1724 -----GACAGCAAGAGTTCGCTCGGTCCCATCAGCCAGAGAAGATGTATGCTGCT 1777
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 155 LeuLysGluThrGluLys----- 160
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1778 TGTGGGCGATCGGGGTGGAGCAGTGACACCTGTGGAATGAGATGCATCTCCAGTGTCT 1837
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 161 -----GlyAspSerValTyrPheAsnAsnAspSerTyrAlaSer 173
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1838 GTGCGATGCGAGTGGGCTGCTGCTGAGGTGCGCTTCAGAGTGGCTGCGAGGAGCTC 1897
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174 AlaArgLeuProCysGlyGlyAlaIleGluAlaCysLysAlaValGluGlyArgVal 193
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1898 AAGAAATGGATTTGCATCATCCGGCCCGCCAGCACCCAGCAATCCACAGCCATG 1957
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 194 LysAsnSerLeuAlaValValArgProGlyHisHisAlaGluProGlnAlaAlaGly 213
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1958 GGATTCGCTTCTCAACTCTGTAGCCATCACCCCAAAA-----CTCTACAGCAGAAG 2011
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 214 GlyPheCysLeuPheSerAsnValAlaAlaLysAsnIleLeuLysAsnTyrPro 233
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2012 TTGAACGTGGCCAGGTCTCTCATCTGAGTGGACATTCACCATGGCATGGCACCAG 2071
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 234 GluSerValArgArgIleMetIleLeuAspTrpAspIleHisHisGlyAsnGlyThrGln 253
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2072 CAGCGGTCTCAATAGCCCTCTGCTCTACATCTCTGATCGCTATGATCAACAGGG 2131
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 254 LysSerPheTyrGlnAspAspGlnValLeuTyrValSerLeuHisArgPheGluMetGly 273
||| ||| ||||| ||||| ||||| ||||| ||||| |||||
```



```

|||||..... ||| |||||
Db 336 AlaLeuValLysLeuLysHisAsnIle-----GluAspValLysArg 349
QY 2594 ATCCAGACCAACACTGAGCTGT 2617
Db 350 ValHisSerLysTyrTrpLysCys 357
RESULT 8
T22134
hypothetical protein F436.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22134
R:Swinsburne, J.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19522
A:Accession: T22134
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-782 <WIL>
A:Cross-references: EMBL:Z50070; NID:e1519046; PIDN:CAA90401.1; GSPDB:GN00020; CESP:F436
A:Experimental source: clone F4366
C:Genetics:
A:Gene: CESP:F4366.4
A:Map position: 2
A:Introns: 81/1; 121/2; 160/1; 204/3; 236/3; 274/3; 324/3; 489/1; 551/2
Alignment Scores:
Pred. No.: 3,32e-18 Length: 782
Score: 486.00 Matches: 172
Percent Similarity: 40.78% Conservative: 100
Best Local Similarity: 25.79% Mismatches: 263
Query Match: 9.13% Indels: 132
DB: 2 Gaps: 19
US-09-502-945-2 (1-2885) x T22134 (1-782)
QY 899 CCGGGGATCGGCCCTGAGCCGACCTAGTCTCACCCTGCGCAGAGTCCCGAGGCC 958
Db 17 ProThrSerProSerAlaValPheProProLeuHisIlePheProHis 36
QY 959 CTCGAG-----CAGCTGTCTATGCAACACAGCAGCAGCAG 994
Db 37 MetGlnGlnTyrPheLeuAsnIleMetGlnLeuValGlnGlnTyrGlnGln 56
QY 995 TTCCTGGAG-----AAGCAGAAGCAGCAGCAG 1021
Db 57 ArgLeuSerSerGlySerProValAsnSerAspGlnAlaSerSerArgSerGluSer 76
QY 1022 CTCAGCTGGGCAAG---ATCCTCACCAAGACAGAGGAGGAGTGCACGAGCCACCCACC 1078
Db 77 ProGlnThrLysLysLeuIleLeuAspGluAspGlyCysAlaThrCysSerValCysThr 96
QY 1079 CAC-----CCTGAG---GAGACAGAGGAGGAGCTGACGAGCAGCAGGAGGTCTTGCTG 1129
Db 97 GluLysValProGluAlaGlnTyrPheSerHisIleGluLeuGluLysGluArgLeuIle 116
QY 1130 GGGGAGGAGCCCTGACCATGCCCCGGAGGGCTCCACACAGAGTGTAGACACACAGGAA 1189
Db 117 SerTyrIleThrLeuThrLysGluLysArg---GluArgAspArgGluAsnThrProAsp 135
QY 1190 GACCTGGAGGAGGAGCAGGAGGAGAG---GATGGGAGGAGGAGGAGGATTCATCCAG 1216
Db 136 HisPheAspGlnLysArgLysArgGluLeuLeuLeuGlnArgIleArgAsnAsnGlnAsn 155
QY 1217 -----GATGGGAGGAGGAGGAGGAGGATTCATCCAG-----GTT 1249
Db 156 LysArgGlnSerLeuLysArgGlyProGlnLeuValArgAspCysLeuThrProPheSer 175
QY 1250 AAGGACGAGGAGGAGGAGGAGTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTGTGTGCT 1309
Db 176 ArgGlnSerAsnAspGluThrGlySerSerGluSerProAspMetLysLysAspAspGlu 195

```

```

QY 1310 GGATACAAAACACTGTTCTCAGATGCCCAACCG----- 1342
Db 196 PheTyrMetLysCysThrThrCysHisGlnProCysSerTyrAlaIleValMetSerAla 215
QY 1343 -----CTGCAACCT 1351
Db 216 PheAspArgProLysCysGlnIleCysPheAspLeuValArgAlaAlaValLeuAsnSer 235
QY 1352 TTGCAGGTGTACCAAGCCCTCAGCCTGGCCACT-----GTGCC 1393
Db 236 AlaSerLeuTyrLysAsnSerLeuLeuLeuPheSerGluLysAsnTyrIleGluIleAsn 255
QY 1394 CACCAAGCCCTGGCGTACCCCAATCTCCCTGCTGCCCTGGGCGCATGAGAACCC 1453
Db 256 AsnSerLeuLeuLysSerIleLysHisSerHisArgLeuMetAsnIleGlnLysAsnMet 275
QY 1454 CCAGACCAACCCGCTCAAGCACCTCTTCCACCAAGTGTGCTTACGACAGTTCATGCTA 1513
Db 276 SerSerLysProPheGlyTyr-----ValTyrAspGluArgMetLeu 289
QY 1514 AAGCAGCAGTGCATGTGCGGGAACACACAGCTGCACCCCTGAGCATGCTGCCGAGTCCAG 1573
Db 290 GlyHisGluCysLysTyrAspSerThrMetAla-----GluCysProLysArgMetLys 307
QY 1574 AGCATCTGGTCCCGCTGCAGGAGACAGGCGCTTAGCAAGTGCAGCGGATCCAGGT 1633
Db 308 LeuIleTyrGluArgLeuGlnLysAspLysLeuLeuAspGlyAlaValLysIleAspAla 327
QY 1634 CGCAAGCCACGCTAGATGAGATCCAGACAGTGCACCTCTGAATACCACACCCCTGCTCTAT 1693
Db 328 ArgGluAlaAsnAspSerGluIleArgLeuAsnHisProGln----- 341
QY 1694 GGGACCATGCTCCCTCAACCGGCGAGAGCTAGACAGCAAGTGTGCTCGGTCCCATCAGC 1753
Db 342 -----GlnLeuIleAspGluIleValSerLeuAsnSerThrGlu 354
QY 1754 CAGAAGATGTATGCTGTGCTGCTTGTGGGGCATCGGGTGGACAGTGCACCGGTGCG 1813
Db 355 Lys-----CysGluGluTyrCysLysAspLysGluIleLeuTrp 367
QY 1814 AATGAGATGACATCTCCAGTGTGTCATGTCAGTGGCTGGCTGCTGTGGAGTGGCC 1873
Db 368 ThrCysGluAsnThrAsnGluAlaAlaArgValAlaIleGlySerIleGluLeuVal 387
QY 1874 TTCAAGGTGCTGCAGGAGCTCAAGAATGATTTGCCATCATCCGCGGCCCCCAGGACAC 1933
Db 388 LysAlaAlaMetGluGlyLysIleHisAsnGlyPheAlaMetValArgProGlyHis 407
QY 1934 CAGCGCAGGAATCCACAGCATGGGATCTGCTTCTTCAACTGTAGCCATCACCGCA 1993
Db 408 HisSerTyrGlyLysValAlaGlnGlyTyrCysIlePheAsnAsnValAlaIleAla 427
QY 1994 AAATCTCTACAGCAAGTGAACGTGGCAAGGTCTCATCTGCTGGACTGGACATTCAC 2053
Db 428 LysTyrAlaIleGluLysLeuGluValLysValAlaIleValAspPheAspTyrHis 447
QY 2054 CATGGCAATGGACCCAGGCGGTCTCAATGACCCCTCTGCTCTACATCTCTCTG 2113
Db 448 AlaGlyAsnGlyThrTyrGlnSerValLysAspGluSerArgIleHisLeuThrSerPhe 467
QY 2114 CATCGTATGACAAACGGAACTTCTTTCAGGCTCTGGGGCTCTGAGAGGTGGTGA 2173
Db 468 HisGlyHisPheGlyAlaPheTrpProPheSer----- 479
QY 2174 GGACAGGCGTGGGTACAATGTGAAC-----GTGCATGGACAGA 2215
Db 480 ---ArgAspTyrAspTyrAlaThrAsnSerGlnAsnThrLeuPheValProLeuAsnGly 498
QY 2216 GGTGTGACCCCTTTCAGGAGCTGGAGTACCTTACAGCCTTCAGACAGCTGTGTATG 2275
Db 499 ThrMetAsn-----SerGluGlyAspTyrValSerValPheHisValLeuLeu 515
QY 2276 CCATTTGCCACAGAGTTCTCACTGTATGTGGTCTAGTCTCCCGGGGTTTTCATGCTGTT 2335

```


A:Cross-references: GB:AE008917; PIDN:AAL52685.1; PID:gl17983511; GSPDB:GN00190
A:Experimental source: strain 16M

C:Genetics:
A:Gene: BME11504
A:Map position: I
C:Keywords: hydrolase

Alignment Scores:

Pred. No.: 1,23e-15 Length: 337
Score: 438.50 Matches: 116
Percent Similarity: 47.84% Conservative: 50
Best Local Similarity: 33.43% Mismatches: 142
Query Match: 8.20% Indels: 39
DB: 2 Gaps: 11

US-09-502-945-2 (1-2885) x AB3440 (1-337)

```
QY 1481 ACCACAGTGTGCTACGACAGTTCATGCTAAAGCACCGAGTCATGTCCGGGAACACA 1540
||||| : : : : : |||||
Db 23 ThrThrArgLeuTyrTrpHisProIleTyrLeuGluHisLeuThrProProGly----- 40

QY 1541 CACGTGCACCTCAGCATGCTGCCGATCCAGAGATCTGCTCCCGGTGCGAGGAGACA 1600
||||| : : : : : |||||
Db 41 -----HisProGluArgProAspArgIleArgAlaLeuMetSerGluLeuGluGlyPro 58

QY 1601 GGCTGTGTAGCAAGTGCAGCGGATCCGAGGTGCGCAAGCCAGCTAGATGAGATCCAG 1660
: : : : : : : : : : : |||||
Db 59 AspPheTyr---ArgLeuAspArgValGluAlaProHisAlaGlyGluAlaAlaLeu 77

QY 1661 ACAGTGCACCTCAATACACACACCTGCTCTATGGGACCATGCTCCCTCAACCGGAGAAG 1720
||||| : : : : : |||||
Db 78 LeuAlaHisProGluGluHisLeu-----GluAlaValArgSerLys 91

QY 1721 CTAGACAGCAAGAAGTGTCTCGTCCATCAGCAGCAAGATGATGCTGTGCTGCTTGT 1780
: : : : : ||||| : : : : :
Db 92 IleProGluProValGluAspGlyGluAlaSerGlnProIle----- 105

QY 1781 GGGGGCATGGGTGCAGTGCACCGTGTGGAATGAGATGCATCTCCAGTGTGCTGTG 1840
: : : : : ||||| : : : : :
Db 106 -----ValLysLeuAspGlyAspThr---TyrValSerProLysSerMetAspAlaAla 122

QY 1841 CGCATGCAGTGGCTGCTGCTGGAGTGGCTTCAAGTGGCTGCAGAGAGAGCTCAAG 1900
||||| : : : : : |||||
Db 123 LeuThrAlaIleGlyAlaAlaMetAlaAlaValAspValMetSerGlyAlaAlaAsp 142

QY 1901 AATGATTTGCCATCATCCGCCCCCAGGACACCGCGAGGAAATCCACGATCCAGGGA 1960
||||| : : : : : ||||| : : : : :
Db 143 AsnValPheValAlaSerArgProProGlyHisHisAlaGluArgSerArgAlaMetGly 162

QY 1961 TTCTGCTTCTTCAACTCTGTAGCCATCACCGCAAACTCTTACAGCAGGAAGTTGAACGTG 2020
||||| : : : : : ||||| : : : : :
Db 163 PheCysValPheAsnAsnIleAlaIleAlaAlaArgHisAlaGlnArgHisGlyLeu 182

QY 2021 GGCAGGCTCTCATCTGAGTGGACATCCACATGGCAATGGCCAGCAGGCGTTTC 2080
: : : : : ||||| : : : : :
Db 183 GluArgIleAlaIleValAspTrpAspValHisHisGlyAsnGlyThrGlnAspIlePhe 202

QY 2081 TACAATGACCCCTCTGCTGTACATCTCTCTCATCGCTATGACACACCGGAACCTCTTT 2140
: : : : : ||||| : : : : :
Db 203 LysAspAspProGlyValMetPheCysSerThrHisGlnPhePro-----LeuTyr 219

QY 2141 CCAGGCTCTGGGCTCTCTGAAGAGGTGTGGTGGAGGACCGGCTGGGTACAAATGTGAAC 2200
||||| : : : : : |||||
Db 220 ProGlySerGlyAspLysHisGluThr-----GlyValGlyAsnIleValAsn 235

QY 2201 GTGGCATGACAGGAGGTGTGACCCGCCATTTGGAGAGCTGGAGTACCTTACAGCCTTC 2260
: : : : : ||||| : : : : :
Db 236 AlaPro-----LeuSerProAsnThrGlySerArgGluPheArgGluAlaPhe 251

QY 2261 AGGACAGTGGTATGCCCATTTGCCACAGATTTCTACCTGATGTGGTCTAGTCTCCGCC 2320
: : : : : ||||| : : : : :
Db 252 AsnSerArgIleLeuProAlaLeuAspAsnPheArgProAspLeuIleLeuSerAla 271
```

```
QY 2321 GGTTTGTATGCTGTGAAGGACATCTG---TCTCCTCTGGGTGGCTACTCTGTCAACCGCC 2377
||||| : : : : : |||||
Db 272 GlyPheAspAla-----HisPheArgAspProLeuAlaGluIleAsnLeuAspGlu 288

QY 2378 AGATGTTTGGCCACTTGACAGGACAGCTGATGACCTCGCA-----GGGGCC 2425
||||| : : : : : |||||
Db 289 SerAspPheAspTrpAlaThrGlyLysLeuMetGluArgAlaGluArgPheCysAspHis 308

QY 2426 CGGTGTGCTGCTGCCCTGAGGAGGCCATGATCTGACCGCCATCTGTGATCCCTCTGAA 2485
||||| : : : : : |||||
Db 309 ArgLeuValSerValLeuGluGlyGlyTyrAspLeuGluGlyLeuSerGlnSerAlaSer 328

QY 2486 GCTTGTGTCTCGGCTCTGCTC 2506
: : : : : |||||
Db 329 ValHisIleThrArgLeuLeu 335
```

RESULT 11

T05998
hypothetical protein F17M5.230 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C:Accession: T05998
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15263
A:Accession: T05998
A:Molecule type: DNA
A:Residues: 1-359 <BEY>
A:Cross-references: EMBL:AL035678; GSPDB:GN00062; ATSP:F17M5.230
A:Experimental source: cultivar Columbia; BAC clone F17M5
C:Genetics:
A:Gene: ATSP:F17M5.230
A:Map position: 4
A:Introns: 12/3; 37/3; 71/3; 103/1; 120/2; 199/3; 260/2
C:Superfamily: acetylcholine aminohydrolase; RPD3/acuc homology

Alignment Scores:

Pred. No.: 2.55e-15 Length: 359
Score: 432.50 Matches: 108
Percent Similarity: 50.45% Conservative: 59
Best Local Similarity: 32.63% Mismatches: 131
Query Match: 8.09% Indels: 33
DB: 2 Gaps: 9

US-09-502-945-2 (1-2885) x T05998 (1-359)

```
QY 1547 CACCTGAGCATGCTGCCGGATCCAGAGCATCTGTCTCCGCTGCAGGAGAGCCCTG 1606
||||| : : : : : |||||
Db 15 HisProGluCysSerAlaArgValProAlaIleValAlaAsnAlaLeuGluMetAsnGluLeu 34

QY 1607 CTAGCAAG-----TGCAGCGGATCCGAGGTGCGAAAGCCACGCCTA 1648
||||| : : : : : |||||
Db 35 ThrProLysPheArgGlySerGlnIleLeuGluAlaAsnPheLysThrAlaThrVal 54

QY 1649 GATGAGATCCAGACAGATGCTGATACACACCTGCTCTATGGACGACGATCCCTC 1708
||||| : : : : : |||||
Db 55 GluAspIleAlaAsnValHisAspLysAlaTyr-----ValPheGlyLeuGluLysAla 72

QY 1709 AACCGGACAGAGTACAGCAGCAAGATGCTGCTCCATCCATCAGCAGAGATGTATGCT 1768
: : : : : ||||| : : : : :
Db 73 MetAspGluAlaSerAspSerGlyLeuIlePheIleGlySerGlyProThrTyrAla 92

QY 1769 GTGCTGCTTGTGGGGCATCGGGGTGGACAGTGCACCGGTGGGAATGAGATGCATCC 1828
: : : : : |||||
Db 93 -----ThrSerThrValThrAlaLeuCysSerSer 102

QY 1829 TCCAGTGTGTCGCATGCGAGTGGCTGCCTGTCTGGAGCTGGCCTTCAAGTGGCTGCA 1888
||||| : : : : : |||||
Db 103 AlaGlyAlaGlyMetAlaLeuValAspSerValValMetSerArgLeuThrIleAsnSer 122

QY 1889 GGAGAGCTCAAGAATGGATTTGCCATCATCCGCCCCCAGGACACCGCCGAGGAATCC 1948
: : : : : ||||| : : : : :
Db 123 ValAspProIleGlyPheAlaLeuIleArgProProGlyHisHisAlaValProLys 142
```


QY 1949 ACAGCATGGATCTCTGCTTCACTCTGTACCATCACCAGAACTCCCTACACGAG 2008
|||||
Db 143 GlyProMetGlyPheCysValPheGlyAsnValAlaIleAlaAraGHisAlaGlnArg 162
|||||
QY 2009 AAGTTGAACGTGGGACGGCTCTCGTGGACATTCACCATGGCAATGGCGACCC 2068
|||||
Db 163 ThrHisGlyLeuLysArgIlePheIleAlaIleAspPheValHisHisGlyAsnGlyThr 182
|||||
QY 2069 CAGCAGCGGTCTTACAATAGCCCTCTGTCTACATCTCTGTCATCGCTATGACAAC 2128
|||||
Db 183 AsnAspAlaPheThrGluAspProAspIlePhePheLeuSerThrHisGln-----Asp 200
|||||
QY 2129 GGGAACCTCTTCCAGGCTCTGGGCTCTCTCAAGAGGTTGGTGGAGGACCGCTGGGG 2188
|||||
Db 201 GlySer-----TyrProGlyThrGlyLysIleSerAspIleGlyLysGlyGlyGluGly 219
|||||
QY 2189 TACAATGTGAACGTGGACATGGAGGAGGTGGACCCGCCCATTTGGAGAGCTGGAGTAC 2248
|||||
Db 220 ThrThrLeuAsnLeuProLeuProGlyGly-----SerGlyAspIleAlaMet 235
|||||
QY 2249 CTTACAGCCTTCAGACAGTGGTGTGNGCCCATTTGCCACGAGTTCTACCTGATGTGTC 2308
|||||
Db 236 ArgThrValPheGluGluIleIleValProCysAlaGlnArgPheLysProAspIleIle 255
|||||
QY 2309 CTAGTCTCGCGCGGCTTGTGATGCTGTTGAAGGACAT---CTGTCTCTCTGGTGGCTAC 2365
|||||
Db 256 LeuValSerAlaGlyTyrAspAla-----HisValLeuAspProLeuAlaAsnLeu 272
|||||
QY 2366 TCTGTCAACCGCAGATGTTTGGCCACTTGACAGGCACTGATGACCTCGGCA----- 2419
|||||
Db 273 GlnPheThrThrAlaThrTyrTyrSerLeuAlaLysAspIleLysArgLeuAlaLysGlu 292
|||||
QY 2420 -----GGGGCGGGTGGTGTGCTGCCCTCGAGGAGGCGCATGACTTACCGCCATCTGT 2473
|||||
Db 293 ValCysGlyGlyArgCysValPhePheLeuGluGlyGlyTyrAsnLeuGluSerLeuSer 312
|||||
QY 2474 GATGCTCTGAAGCTGTGTCTCGCTCTGCTC 2506
|||||
Db 313 SerSerValAlaAspSerPheArgAlaLeuLeu 323
|||||

RESULT 12
D97450
histone deacetylase family protein nmb0605 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: D97450
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: D97450
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-311 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK86557.1; PID:g15155719; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C1358
A:Map position: circular chromosome
C:Superfamily: acetylpolymine aminohydrolase; RPD3/acuc homology

Alignment Scores:
Pred. No.: 1,14e-14 Length: 311
Score: 420.00 Matches: 107
Percent Similarity: 49.68% Conservative: 50
Best Local Similarity: 33.86% Mismatches: 121
Query Match: 7.86% Indels: 38
DB: 2 Gaps: 10

US-09-502-945-2 (1-2885) x D97450 (1-311)

QY 1547 CACCTTGAGCATCTGGCCGGATCCAGAGCATCTGTGCCGGTGTGACGAGACAGCCCTG 1606

Db 20 HisProGluArgProAspArgLeuArgSerLeuAsnIleAlaLeuGluHisProAsn--- 38
|||||
QY 1607 CTTAGCAAGTGCAGGCGGATCCAGGTCCGAAGCCACGCTAGATGAGATCCAGACAGTG 1666
|||||
Db 39 PheGluArgLeuAspArgLysGluAlaProGlnAlaAsnGluAspAlaValLeuLeuAla 58
|||||
QY 1667 CACTCTGAATACCACACCCCTGCTCTATGGACCCAGTCCCTCAACCGGAGAGAGCTAGAC 1726
|||||
Db 59 HisProGluGluHisLeuLeu-----AlaValMetArgGlnValProGlu 73
|||||
QY 1727 AGCAAGAAGTGTGCTGGTCCCATCACCACAGAGATGTATGCTGTGCTCTGCTGGGGGC 1786
|||||
Db 74 GluAsp-----GlyGluIleAsnArg----- 80
|||||
QY 1787 ATCGGGGTGACAGTGCACACCGTGTGAATGAGATGACATCTCCAGTGTGTCGCGCATG 1846
|||||
Db 81 -----IleGluSerAspThr---TyrLeuSerProLysSerLeuGlnAlaAlaLeuThr 97
|||||
QY 1847 GCAGTGGGCTGCTGCTGGAGCTGGCTTCAAGTGTGGCTGCAGGAGAGCTCAAGAATGGA 1906
|||||
Db 98 GlyIleGlyAlaAlaMetAlaAlaValAspValPheSerGlyAlaAlaAspAsnVal 117
|||||
QY 1907 TTTGCCATCATCCGGCCCCCAGGACACCCAGCGAGGAATCCACAGCATGGATTCGTC 1966
|||||
Db 118 PheValAlaAlaArgProProGlyHisHisAlaGluThrAlaLysAlaMetGlyPheCys 137
|||||
QY 1967 TTTCTCAACTCTGTAGCCATCACCAGCAAACTCCTACAGCAGAGTTGAACGTGGGCAAG 2036
|||||
Db 138 LeuPheAsnAsnValAlaIleAlaAlaArgHisAlaGlnLysAlaHisGlyAlaGluArg 157
|||||
QY 2027 GTCTCATCTGTGGACTGGGACATTCACCATGGCAATGGCACCCAGCGAGGCTTCTACAAT 2086
|||||
Db 158 IleAlaIleIleAspTrpAspValHisHisGlyAsnGlyThrGlnAspIlePheTrpAsn 177
|||||
QY 2087 GACCCCTGTGCTCTACATCTCTCTCATCGCTATGACACAGGGAATCTCTTCTCCAGGC 2146
|||||
Db 178 AspIleSerValLeuPheCysSerThrHisGlnMetPro-----LeuTyrProTrp 194
|||||
QY 2147 TCTGGGCTCCTGAAGAGTTGTGGAGGACCGAGCTGGGTACAATGTGAACGTGGCA 2206
|||||
Db 195 SerGlyAspLysAsnGluThr-----GlyValLysAsnAsnIleValAsnAla 210
|||||
QY 2207 TGGACAGAGGTGTGGACCCGCCCATTTGGAGACGTGGAGTACCTTACAGCTTCAGGACA 2266
|||||
Db 211 -----ProLeuSerProAsnThrGlySerAspHisPheArgGluAlaPheLysSer 227
|||||
QY 2267 GTGTGTATGCCATTCGCCAGGATCTCTACCTCATGTGTGCTCTAGTCTCCGCGGGTGT 2326
|||||
Db 228 ArgValLeuProAlaIleAlaAspPheSerProAspLeuIleLeuIleSerAlaGlyPhe 247
|||||
QY 2327 GATGCTGTTGAAGGACATCTCTCTCTGGTGGTCTGTCTGTCCCGCCAGATGTTTT 2386
|||||
Db 248 AspAla-----HisHisArgAspProLeuAlaGlnIleAsnLeuValGlyLysAspPhe 265
|||||
QY 2387 GGCCACTTGCAGGACAGCTGATGACCCCTG-----GCAGGGGCGCGGTGGT 2434
|||||
Db 266 AspTrpAlaThrGlyArgLeuLeuGluMetAlaAspLysTyrAlaSerAsnArgValVal 285
|||||
QY 2435 CTGGCCCTGGAGGAGGCGCATGCTGACCCCATCTGTGTATGCTCT 2482
|||||
Db 286 SerLeuLeuGluGlyGlyTyrAspLeuGluGlyLeuAlaGluSerAla 301
|||||

RESULT 13

AF2668

deacetylase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002

C:Accession: AF2668

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AF2668
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-311 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAL41764.1; PID:gl7739116; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu0748
 A:Map position: circular chromosome
 C:Superfamily: acetylpoliamine aminohydrolase; RPD3/acuc homology

Alignment Scores:
 Pred. No.: 1.14e-14 Length: 311
 Score: 420.00 Matches: 107
 Percent Similarity: 49.68% Conservative: 50
 Best Local Similarity: 33.86% Mismatches: 121
 Query Match: 7.86% Indels: 38
 DB: 2 Gaps: 10

US-09-502-945-2 (1-2885) x AF2668 (1-311)

QY 1547 CACCTGAGCATGTCGCGGATCCAGCATCTGTCGCGCTGCAGAGACAGCGCTG 1606
 Db 20 HisProGluArgProAspArgLeuArgSerLeuAsnIleAlaLeuGluHisProAsn--- 38
 QY 1607 CTTAGCAAGTCGCGAGCGATCCGAGGTCGCAAGCCAGCTAGATGATCCAGACAGT 1666
 Db 39 PheGluArgLeuAspArgLysGluAlaProGlnAlaAsnGluAspAlaValLeuAla 58
 QY 1667 CACTCTGAATACACACCCCTCTCTATGGGACCATCCCTCAACCGCGAGAGCTAGAC 1726
 Db 59 HisProGluHisLeuLeu-----AlaValMetArgGlnValProGlu 73
 QY 1727 AGCAAGAAGTTGCTGCTCCATCAGCAGCAGAGATGATGCTGCTGCTGCGGGC 1786
 Db 74 GluAsp-----GlyGluLeuAsnArg----- 80
 QY 1787 ATCGGGGTGACAGTACACCGTGTGGATGAGATGATGATGATGATGATGATGATG 1846
 Db 81 -----IleGluSerAspThr-----TyrLeuSerProLysSerLeuGlnAlaAlaLeuThr 97
 QY 1847 GCAGTGGCTGCTGCTGAGCTGGCTTCAAGTGGCTGCTGCTGCTGCTGCTGCTGCT 1906
 Db 98 GlyIleGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 117
 QY 1907 TTTGCCATCATCGCGCCCGCCAGGACACCGCGGAGGATCCACAGCATGGGATTTCTGC 1966
 Db 118 PheValAlaAlaAlaArgProProGlyHisHisAlaGluThrAlaLysAlaMetGlyPheCys 137
 QY 1967 TTTCTCACTCTGATGCTACCGCAGCGGAACTCCTACAGCAGAGTGAACGTGGCAAG 2026
 Db 138 LeuPheAsnAsnValAlaIleAlaAlaArgHisAlaGlnLysAlaHisGlyAlaGluArg 157
 QY 2027 GTCTCATCTGCTGAGCATTCACCATGCGCAATGGCAGCCAGCCAGCGGCTTCAAT 2086
 Db 158 IleAlaIleAspTrpAspValHisHisGlyAsnGlyThrGlnAspIlePheTrpAsn 177
 QY 2087 GACCCCTCTGCTGTACATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2146
 Db 178 AspIleSerValLeuPheCysSerThrHisGlnMetPro-----LeuTyrProTrp 194
 QY 2147 TCTGGGCTCTCCTGAAGAGTTGCTGGAGGACCGAGGCTGGGTGATGATGATGATGAT 2206
 Db 195 SerGlyAspLysAsnGluThr-----GlyValLysAsnAsnIleValAsnAla 210
 QY 2207 TGGACAGGAGGTGTGACCCCGCCCATTTGAGAGCTGGAGCTGATACCTTACAGCTTCA 2266
 Db 211 -----ProLeuSerProAsnThrGlySerAspHisPheArgGluAlaPheLysSer 227

QY 2267 GTGCTGATGCCATTTGCCACGAGTTCTACCTGATCTGCTGCTGCTGCTGCTGCTT 2326
 Db 228 ArgValLeuProAlaIleAlaAspPheSerProAspLeuIleLeuIleSerAlaGlyPhe 247
 QY 2327 GATGCTGTTGAAGGACATCTGCTCTCTGCTGGTGGCTACTCTGTCACCGCCAGATGTTT 2386
 Db 248 AspAla-----HisHisArgAspProLeuAlaGlnIleAsnLeuValGlyGluAspPhe 265
 QY 2387 GGCACACTGACAGGAGCGAGCTGATGACCCCTG-----GCAGGGGCGCGGTGGT 2434
 Db 266 AspTrpAlaThrGlyArgLeuLeuGluMetAlaAspLysTyrAlaSerAsnArgValVal 285
 QY 2435 CTGGCCCTGAGGAGGCGCATGACTGACCGCCATCTGTCGCTGCTGCTGCTGCT 2482
 Db 286 SerLeuLeuGluGlyGlyTyrAspLeuGluGlyLeuAlaGluSerAla 301

RESULT 14

T27101

hypothetical protein Y51H1A.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27101

R:Smyle, R.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20309

A:Accession: T27101

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-517 <WIL>

A:Cross-references: EMBL:AL032644; PIDN:CAA21669.2; GSPDB:GN00020; CESP:Y51H1A.5

A:Experimental source: clone Y51H1A

C:Genetics:

A:Gene: CESP:Y51H1A.5

A:Map position: 2

A:Introns: 41/3; 84/3; 218/1; 408/1

Alignment Scores:

Pred. No.:	5.98e-14	Length:	517
Score:	406.50	Matches:	127
Percent Similarity:	47.48%	Conservative:	71
Best Local Similarity:	30.46%	Mismatches:	152
Query Match:	7.60%	Indels:	67
DB:	2	Gaps:	17

US-09-502-945-2 (1-2885) x T27101 (1-517)

QY 1490 GTGGTCTACGACACGCTTCATGCTAAAGCACCAGTCATGTGCGGGAACACACACGTCGAC 1549
 Db 9 ValIlePheAspHisGlnGluArgHisAspGlnProTrpProSerTyrHisIle--- 27
 QY 1550 CTTGACGATGCTGGCGGATCCAGAGCATCTGTCCTGGCTGCAGGACAGGCTGCTT 1609
 Db 28 ---GluValProArgArgLeuAspAlaIleLeuGluArgLeuAsnThrThrLysLeuLeu 46
 QY 1610 AGC-----AAGTGGCGCGGATCCGAGTCCCAAGCCACCATAGATGAGATCCAGACA 1663
 Db 47 ThrAspProArgIleGluHisIleProArgGluAlaGluGluSerGluIleLeuAla 66
 QY 1664 GTGCACTCTGAATACACACCCCTGCTCTATGGACCATGTCCTCAACCGCGGAGAAGTA 1723
 Db 67 ValHisThrLysArgTyrVal-----AspAspVal 76
 QY 1724 GACACAGAAGATGCTCGTCCCATCAGCCAGAGAAGATGATGCTGCTGCTGCTGCTGGG 1783
 Db 77 LysSerThrGluThrMetThrValGluGlnGlnGluSerPheCys----- 91
 QY 1784 GCATCGGGTGCAGAGTGCACACCGTGTGGAATGAGATGACTCTCCCTCCAGCTGCT 1837
 Db 92 -----ThrLysTyrGluAspIleTyrValAsnSerAlaThrTrp 104
 QY 1838 -----GTGCGCATGGCAGTGGGCTGCTGCTGGAGCTGGCCTTCAAGCTGCTGACAGA 1891
 Db 105 HisArgAlaLysLeuAlaAlaGlyAlaSerIleAspLeuMetThrSerValMetAlaAla 124

A;Cross-references: GB:AE004796; GB:AE004091; NID:g9949939; PIDN:AAG07161.1; GSPDB:GN
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA3774

Alignment Scores:		
Pred. No.:	4,84e-13	Length:
Score:	389.00	Matches:
Percent Similarity:	46.72%	Conservative:
Best Local Similarity:	29.06%	Mismatches:
Query Match:	7.28%	Indels:
DA:	2	Gaps:
		9
		58
		129
		62
		102
		380

US-09-502-945-2 (1-2885) x D83174 (1-380)

Qy	1550	CCTGAGCATCTGGCCGCATCCAGACATCTGTGCCCGCTGCAGGAGACAGCGCTGCTT	1609
Db	43	ProgluThrLysArgLeuLysSerLeu-----LeuAspValSerGlyLeuThr	59
Qy	1610	AGCAAGTGCGAGCGGATCCGAGGTGCRAAAGCCACGCTAGATGATGATCCAGACAGTCGAC	1669
Db	60	AlaArgLeuGlnLeuArgSerAlaProProAlaSerAspGluAspLeuLeuArgValHis	79
Qy	1670	TCTGAATACCACACCCTGTCTCTATGGCACAGTCCCTC AACCGGCAGAAGCTAGACAGC	1729
Db	80	ProAlaHisTyrLeu-----	84
Qy	1730	AAGAAGTTGCTCGGTCCATCCACGACAAGATGTA TGCTGTGCTGTGGGGGCATC	1789
Db	85	-----GluArgPheLysAlaLeuSerAspAlaGlyGlyGly	96
Qy	1790	GGGTGGACAGTCACACCGGTGTGGAATGAGATGC ACTCCTCCAGTGTGTCGCATGGCA	1849
Db	97	SerLeuGlyGlnAspAlaProIleGlyPro---GlySerTyrGluIleAlaArgLeuSer	115
Qy	1850	GTGGGCTCGCTGCTGAGCTGGCGCTTCAAAGTGGCTGCAGGAGAGCTCAAGAATGGATT	1909
Db	116	AlaGlyLeuAlaIleAlaAlaLeuaspAlaValLeuAlaGlyGluAlaAspAsnAlaTyr	135
Qy	1910	GCCATCATCCGGCCCCAGACACACCGCGAGGAATCCACAGCCATCGGGATCTCGCTTC	1969
Db	136	SerLeuSerArgProProGlyHisHisCysLeuProAspGlnAlaMetGlyPheCysPhe	155
Qy	1970	TTCAACTCTGTAGCCATCACCGCAAAACCTCTACAGCAGAAGTTGAAGCTGGCGAAGTC	2029
Db	156	PheAlaAsnIleAlaValAlaIleGluAlaAlaLysAlaArgHisGlyValcLuArgVal	175
Qy	2030	CTCATCTGTGACCTGGGACATTCACATGCGATGGCATGGCCACCCAGCAGGCGTTTACAATGAC	2089
Db	176	AlaValLeuAspTrpAspValHisHisGlyAsnGlyThrGlnAlaIleTyrTyrArgArg	195
Qy	2090	CCCTCTGTCTACATCTCTCTGCATCCCTATGACACGGGAACCTCTTCCCAGGCGTCT	2149
Db	196	AspAspValLeuSerIleSerLeuHisGln-----AspGlyCysPheProGlyTyr	213
Qy	2150	GGGCTCCTCAAGAGTTGCTGGAGCACGCGCGGTGACATPGTAGCGTGGCATGG	2209
Db	214	SerGlyAlaGluAspIleGlyGluAspArgGlyArgGlyPheAsnLeuAsnValPro---	232
Qy	2210	ACAGAGGTGTGNACCCCCCATTTGAGACGTGGAGTACCTTACAGCCTTCAGGACAGTG	2269
Db	233	-----LeuLeuProGlyGlyGlyHisAspAlaTyrMetGlnAlaMetGlnArgIle	249
Qy	2270	GTGATGCCCATTCGCCACGAGTTCTCACCTGATGTGGTCTAGTCCGCGCGGTTTTGAT	2329
Db	250	ValLeuProAlaLeuGluArgPheArgProGlnLeuIleValAlaIleSerGlyPheAsp	269
Qy	2330	GCTGTGAAGGACATCTCTCTCTCGTGGGTGGCTACTGTGCACCGCCAGATGTTTGGC	2389
Db	270	Ala-----AsnAlaValAspProLeuAlaArgMetGlnLeuHisSerAspSerPheArg	287
Qy	2390	CAC TTGACCGAGCGCTATGATGCCCTG-----GCAGGGGCGCGGTGGTGGTCTG	2437

Qy	1892	GAGCTCAAGATGAGTTCCCATCATCCGGCCCCCAGACACCAACGCGGAGGAATCCACA	1951
Dy	125	Lys---ArgProGlyIleAlaPheIleArgProProGlyHisHisAlaMetProAspGlu	143
Qy	1952	GCCATGGGATTCGCTTCCAACTCTGTAGCCATCACCGCAAAACTCCTACACGACAAG	2011
Dy	144	GlyCysGlyPheCysIlePheAsnAsnValAlaIleAlaAlaLysAlaAlaIleGln---	162
Qy	2012	TTGAACCTGGCGAAGTCTCATCTGCTGGACTGGGACATTCACCATGGCAATGCCACCACG	2071
Dy	163	--AsnGlyGlnLysValLeuIleValAspTyrAspValHisAlaGlyAsnGlyThrGln	181
Qy	2072	CAGCGCTC-----TACAATGACCCTCTGCTCTACATCTCTCTGCATCGCATGAC	2125
Dy	182	GluCysValGluGlnMetGlyGluGlyAsnValGlnLeuIleSerIleHisArgTyrGlu	201
Qy	2126	AACGGGAACCTTCTCCAGGCTCTGGGGCTCTGAAGAGTTGGTGGAGGACCAGGCGTG	2185
Dy	202	AsnGlyHisPhePro-----AsnMetProGln-----ThrGlyIle	214
Qy	2186	GGGTACAATGTGAACGTGCATGGACAGGAGGTGGACCCCCCATTTGGA-----	2236
Dy	215	TyrHisAsn-----TyrLysAsnThrIleAsnLeuProLeuAsnThrIleGly	230
Qy	2237	-----CACCTGGAGTACCTTCACAGCCTTCAGGACAGTGGTCATGCCCATTCGCCACGAG	2290
Dy	231	LeuThrAspAlaAspTyrHisAlaLeuPheThrHisIleLeuProThrIleHisAla	250
Qy	2291	TTCTCAACTGATGTGGTCTAGTCTCCGCGGGGTTGATGTGTGTGAAGGACATCTGTCT	2350
Dy	251	PheGlnProAspLeuLeuValSerSerGlyPheAspAlaSerIleGly-----Asp	268
Qy	2351	CCTCTGGGTGGTACTCTGTACCGCCAGATGTTTGGCCACTTGCACGAGCAGCTGATG	2410
Dy	269	ProGluGlySerMetGlnValThrProAlaGlyPheAlaThrMetIleArgMetLeuIle	288
Qy	2411	ACCTGTCAGCGGGCGGGTGCTGGCCCTGGAGGAGGCCATGACTTGCAGCGCCATC	2470
Dy	289	Asp---ThrGlyIleProValAlaAlaLeuLeuGluGlyGlyTyrPheLeuAspAlaLeu	307
Qy	2471	TGTGATCCCTCTGAAGCTTGCTGTCTCGGCTCTGCTCAGT-----GTAAAG	2515
Dy	308	AlaAlaAspSerGluTrpValLeuArgAlaLeuLeuGlyGluGluIleProArgIleArg	327
Qy	2516	CTGACGCCCTTGGATAGGCACTCTTGCAGCAAAAGCCCCAACATCAACGCACTGCCACG	2575
Dy	328	ValGluLysIleHisSerAlaIle-----AlaAspThrIleGlyArg	341
Qy	2576	CTAGAGAAAGTCATPCGAGATCCAGACGAACACTGTGGAGCTGTGTGCACAAGTTGCCGCT	2635
Dy	342	ValValLysArgTyrGluGlySerCysProPhePheLysLysValGlnGluLeuArgGly	361
Qy	2636	GGTCTGGCGGGTCCCTCGCGAGGGGGCCCAAGCAGGTGAGCCGAAGAAGCC	2686
Dy	362	IleLeuGly-----CysArgValAlaGluGluAspGlnGlnGluAla	376

RESULT 15

DB3174
probable acetylpolylamine aminohydrolase PA3774 [imported] - Pseudomonas aeruginosa (strain
C;Species: pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: DB3174
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Miziochuchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: AB2950; MUID:2043737; PMID:10984043
A;Accession: DB3174
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-380 <STO>

Db 288 AlaMetThrAlaMetValArgAspAlaAlaGluArgHisAlaGlyGlyArgLeuValVal 307
QY 2438 GCCCTGGAGGAGGCCCATGACTTGACCGCCATCTGTGATGCTCTCTGAAGCTTGTGTC--- 2494
|||||
Db 308 ValHisGluGlyGlyTyr-----SerGluAlaTyrValPro 319
QY 2495 -----TCGGCTCTGCTCAGTGTAAAGCTGCAGCCC 2524
||| :|||
Db 320 PheCysGlyLeuAlaValIleGluGluLeuSerGlyValArgSerAlaValArgAspPro 339
QY 2525 TTGATGAGGCAGTCTTGCAGCAAAAGCCCAAC 2557
||| :|||
Db 340 LeuArgAspPheIleGluLeuGlnGlnProAsn 350

Search completed: March 21, 2003, 13:04:54
Job time : 78.3908 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 13:00:02 ; Search time 30.8914 Seconds

(without alignments)
9985.861 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346

Sequence: 1 ggaattctctgttcgaagt.....aaacaaaagttaaaattt 2885.

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 442306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09502945/runat_14032003_101101_19301/app_query.fasta_1.10979
-DB=PublishedApplications-AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -MODE=LOCAL -OUTFMT=ptc -THR_SCORE=ptc -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09502945 -CGN_1_1_60_@runat_14032003_101101_19301
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:
2: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pcp:
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp:
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:
7: /cgn2_6/ptodata/2/pubpaa/PTCTUS_PUBCOMB.pcp:
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp:
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp:
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp:
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp:
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pcp:
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4793	89.7	1122	9	US-10-072-094-8
2	4793	89.7	1122	9	US-10-072-094-91
3	3612.5	67.6	717	10	US-09-817-913-9
4	3612.5	67.6	717	10	US-09-817-538-9

5	2691	50.3	1084	9	US-10-072-094-7
6	2685	50.2	967	10	US-09-817-913-7
7	2685	50.2	967	10	US-09-817-538-7
8	2686	49.9	1069	9	US-10-072-094-87
9	2579	48.2	1011	9	US-10-072-094-89
10	2212.5	41.4	780	9	US-10-072-094-93
11	2184.5	40.9	1141	9	US-10-072-094-107
12	1988	37.2	879	9	US-10-072-094-90
13	1674.5	31.3	855	9	US-10-072-094-9
14	1671.5	31.3	855	10	US-09-817-913-13
15	1671.5	31.3	855	10	US-09-817-538-13
16	961.5	18.0	590	9	US-10-072-094-99
17	893.5	16.7	495	9	US-10-072-094-95
18	771.5	14.4	1215	10	US-09-976-280A-4
19	771.5	14.4	1215	10	US-09-976-280A-22
20	766.5	14.3	1215	10	US-09-817-913-11
21	766.5	14.3	1215	10	US-09-817-538-11
22	723	13.5	163	9	US-10-072-094-4
23	705.5	13.2	687	9	US-09-976-280A-3
24	672	12.6	706	9	US-10-072-094-6
25	623.5	11.7	673	10	US-09-976-280A-2
26	595	11.1	342	9	US-09-911-150-4
27	543	10.2	175	9	US-10-072-094-5
28	445	8.3	112	9	US-10-072-094-21
29	413	7.7	105	9	US-10-072-094-2
30	374.5	7.0	335	10	US-09-971-309-80
31	316.5	5.9	480	10	US-09-347-331-9
32	306.5	5.7	488	10	US-09-925-300-1563
33	296.5	5.5	150	9	US-09-911-150-6
34	291	5.4	481	10	US-09-817-913-1
35	291	5.4	481	10	US-09-817-538-1
36	289	5.4	576	10	US-09-347-331-11
37	284.5	5.3	2530	9	US-10-184-644-155
38	279.5	5.2	377	10	US-09-784-877-2
39	279.5	5.2	377	10	US-09-817-913-15
40	279.5	5.2	377	10	US-09-817-538-15
41	278.5	5.2	2457	9	US-10-184-644-569
42	278	5.2	517	10	US-09-883-720-8
43	277.5	5.2	458	10	US-09-883-720-2
44	277.5	5.1	1903	9	US-10-184-644-455
45	277	5.1	1815	9	US-10-184-644-19

ALIGNMENTS

RESULT 1
US-10-072-094-8
; Sequence 8, Application US/10072094
; Publication No. US20030025538A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, DONALD
; APPLICANT: LORENZI, MATTHEW
; APPLICANT: ATTAR, RICARDO
; APPLICANT: GOTTARDIS, MARCO
; TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
; FILE REFERENCE: 3053-4145U1
; CURRENT APPLICATION NUMBER: US/10/072, 094
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,296
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-094-8

Alignment Scores:
Pred. No.: 7.84e-228
Score: 4793.00
Percent Similarity: 99.57%
Best Local Similarity: 99.25%
Matches: 1122
Conservative: 3
Mismatches: 4

Query Match: 89.66% Indels: 1
DB: 9 Gaps: 0

US-09-502-945-2 (1-2885) x US-10-072-094-8 (1-1122)

```
QY 2 GAATTCCTCTTGTCAAGTCAAGAGCCACACAGCGGCGCTCAACCATTCCTCCCA 61
DB 189 GluPheLeuLeuSerLysSerLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 208
QY 62 CAGCACCCCAATGCTGGGAGCCACCATGCTCTTGTGGACAGATGCCCTCCCGAG 121
DB 209 GlnHisProLysCysTrpGlyAlaHisAlaSerLeuAspGlnSerSerProGln 228
QY 122 AGCGGCGCCCTGGGACGCTCCCTCCACAACTGCCTTTCCCTGGGCGCTTACACAGT 181
DB 229 SerGlyProGlyThrProProSerTyrLysLeuProLeuProGlyProTyrAspSer 248
QY 182 CGAGAGGACTTCCCTCCGGAACACAGCCTCTGACCCCACTTGAAGTGGCTCAAGG 241
DB 249 ArgAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg 268
QY 242 CTAATAACAGAAGTGTGAGCGGAGAGCAGTCCCTCTCGTCGCAAGGATGGACT 301
DB 269 LeuLysGlnLysValAlaGluArgSerSerProLeuLeuArgLysAspGlyThr 288
QY 302 GTATTAGCACCTTTAAGAGAGAGCTGTTGAGATCACAGGTGCCGGCGCTCGCGCTG 361
DB 289 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyAlaGlyProGlyAlaSer 308
QY 362 TCGGTGTAAACAGCGCCCGGCTCCGCGCCAGCTCTCCCAACAGCTCCACAGCACC 421
DB 309 SerValCysAsnSerAlaProGlySerGlyProSerSerProAsnSerHisSerThr 328
QY 422 ATCGCTGAGAATGGCTTTACTGTGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCAG 481
DB 329 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGln 348
QY 482 CACCGAGCCCTCCCTCTGGACAGCTCCCAACCAAGTCTCAGGCTCTACAGCTCTCCTCT 541
DB 349 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTyrThrSerProSer 368
QY 542 CTGCCCCAATCTCCCTAGGCTGACGCGCAGCGTCACTGTACCAACTCACACCTCACT 601
DB 369 LeuProAsnIleSerLeuGlyLeuGlnAlaThrValThrValThrAsnSerHisLeuThr 388
QY 602 GCCTCCCGCAAGCTGTCACACAGAGAGGCGCGAGAGCAGCGCCCTCCAGTCCCTGCGG 661
DB 389 AlaSerProLysLeuSerThrGlnGlnGluAlaGluArgGlnAlaLeuGlnSerLeuArg 408
QY 662 CAGGTTGGCAGCGTGACCGCAAGTTTCATGAGCACATCCTCTATTCTGGCTGCCCTGCTG 721
DB 409 GlnGlyGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCysLeuLeu 428
QY 722 GCGGTGGCACTGGAGGGCGACGGGACCGCCCGCAGGCGATGCCCTCCCTGTGACGATGG 781
DB 429 GlyValAlaLeuGluGlyAspGlySerProHisGlyHisAlaSerLeuLeuGlnHisVal 448
QY 782 CTGTTCTGGACAGCGCCCGGACAGACAGCACCCTCATTTGCTGTGCACTCCACGGCGAG 841
DB 449 LeuLeuLeuGluGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 468
QY 842 TCCCCACTAGTACGGGTGAACGTGTGGCCACCAGCATCGGAGGTAGGCAAGCTCCCG 901
DB 469 SerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLysLeuPro 488
QY 902 CGGCATCGGCCCTGAGCGCACTCACTAGTCTCACCCTGCGCGAGAGTCCCCAGGCCCTG 961
DB 489 ArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGlnAlaLeu 508
QY 962 CAGCAGCTGTGTATGAACAACAGCACCAGCTTCTCTGGAGAGCAGAGCAGCAG 1021
DB 509 GlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnLysGlnGlnGln 528
```

```
QY 1022 CTACAGCTGGSCAAAGATCCTCACCAAGACAGAGGGAGGTGCCAGGACCCACACCCAC 1081
DB 529 LeuGlnLeuGlyLysIleLeuThrLysThrGlyGluLeuProArgGlnProThrThrHis 548
QY 1082 CCTGAGAGACAGAGGAGGAGCTACCGAGCAGCAGAGGTCTTGTCTGGGGGGAGGCC 1141
DB 549 ProGluGluThrGluGluLeuThrGluGlnGlnGluValLeuLeuGlyGluGlyAla 568
QY 1142 CTGACCATTGCCCGGAGGCTCCACAGAGAGTGGAGAGCACACAGAGAGCTCGAGGAG 1201
DB 569 LeuThrMetProArgGluGlySerThrGluSerGluSerThrGlnGluAspLeuGlu 588
QY 1202 GAGGACGAGGAGGAGTGGGAGGAGGAGGATTGCATCCAGGTTAAGGACGAGGAG 1261
DB 589 GluAspGluGluGluAspGlyGluGluAspCysIleGlnValLysAspGluGlu 608
QY 1262 GGCAGAGTGTGTGTGAGGAGGCGCCGACTTGAGGAGCCTGTGTGTGGATACAAAAA 1321
DB 609 GlyGluSerGlyAlaGluGluGluProAspLeuGluGluProGlyAlaGlyTyrLysLys 628
QY 1322 CTGTTCTCAGATGCCCAACCGCTGCACCTTTGAGGTGTACCAAGCGCCCTCAGGCTG 1381
DB 629 LeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyrGlnAlaProLeuSerLeu 648
QY 1382 GCCACTGTGCCCAACAGCCCTGGGCGGTACCCAAATCCTCCCTGCTGCCCTGGGGGC 1441
DB 649 AlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaProGlyGly 668
QY 1442 ATGAAGAACCCCGACAGCAACCGCTCAAGCACCTTTCACCAACAGTGTGTCTACAC 1501
DB 669 MetLysSerProProAspGlnProValLysHisLeuPheThrThrGlyValValTyrAsp 688
QY 1502 ACGTTCATGCTAAAGCACACAGTGTGCGGGACACACACAGTGCACCCCTGAGCATGCT 1561
DB 689 ThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGluHisAla 708
QY 1562 GCGCGGATCCAGAGCATCTGTGTCGGGCTGAGGAGACAGGCGCTCTTAGCAAGTGCAG 1621
DB 709 GlyArgIleGlnSerIleTrpSerArgLeuGlnGluThrGlyLeuLeuSerLysCysGlu 728
QY 1622 CGGATCCGAGGTGCGAAAGCCAGCTAGATGAGATCCAGAGTGCACCTCTGAAATACCAC 1681
DB 729 ArgIleArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGluTyrHis 748
QY 1682 ACCCTGCTCTATGGACACAGTCCCTCAACCGGCAAGAGCTAGACAGCAAGAAGTTGCTC 1741
DB 749 ThrLeuLeuTyrGlyThrSerProLeuAsnArgGlnLysLeuAspSerLysLysLeuLeu 768
QY 1742 GGTCCCATCAGCCAGAAGATGTATGCTGTGTCCTTGTGGGGCATCGGGGTGACAGT 1801
DB 769 GlyProIleSerGlnLysMetTyrAlaValLeuProCysGlyGlyIleGlyValAspSer 788
QY 1802 GACACCGTGTGGAATGAGATGCACCTCCCTCCAGTGTGTCGATGCGCATGCGCTGGCTG 1861
DB 789 AspThrValTrpAsnGluMetHisSerSerSerAlaValArgMetAlaValGlyCysLeu 808
QY 1862 CTGGAGCTGGCCTTCAAGTGGCTGCAGGAGAGCTCAAGAATGGATTGGCCATCATCCGG 1921
DB 809 LeuGluLeuAlaPheLysValAlaAlaGlyGluLeuLysAsnGlyPheAlaIleIleArg 828
QY 1922 CCCCCAGGACACACCGCCGAGGAATCCACAGCCATGGGATTTCTTCAACTCTGTA 1981
DB 829 ProProGlyHisHisAlaGluSerThrAlaMetGlyPheCysPheAsnSerVal 848
QY 1982 GCCATCCGCAAAACTCCTACAGCAGAGTGAACGTGGGCAAGGTCTCATCTGTCGAC 2041
DB 849 AlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeuIleValAsp 868
QY 2042 TGGGACATTCACCATGGCAATGGCACCGAGCGGCTTCTACAAATGACCCCTCTGTGCTC 2101
DB 869 TrpAspIleHisHisGlyAsnGlyThrGlnGlnAlaPheTyrAsnAspProSerValLeu 888
QY 2102 TACATCTCTCTCATCGCTATGACAAACGGGAACCTTTTCCAGGCTCTGGGGCTCTCTGAA 2161
```


Db 889 TyrIleSerLeuHisArgTyrAspAsnGlyAsnPhePheProGlySerGlyAlaProGlu 908
Qy 2162 GAGTTGTGTGGAGGACCGGTGGGTACATGTGACATGTGACATGTGACATGTGACATGTG 2221
Db 909 GluValGlyGlyGlyProGlyValGlyTyrAsnValAsnValAlaIleThrPheGlyGlyVal 928
Qy 2222 GACCCCCCATTTGGAGAGCTGGAGTACCTTACAGACCTTTGAGACAGTGGTGTGATGCCATT 2281
Db 929 AspProIleGlyAspValGlyTyrLeuThrAlaPheArgThrValValMetProIle 948
Qy 2282 GCCACGAGTTCTCACCTGATGTGTCTAGTCTCCCGCCGCTTTGATGCTGTGTAAGGA 2341
Db 949 AlaHisGluPheSerProAspValValLeuValSerAlaGlyPheAspAlaValGluGly 968
Qy 2342 CATCTGTCTCTCTGGGTGGTACTCTCTACCGCCAGATGTTTGGCCACTTGCACAGG 2401
Db 969 HisLeuSerProLeuGlyGlyTyrSerValThrAlaArgCysPheGlyHisLeuThrArg 988
Qy 2402 CAGCTGATGACCTGGCAGGGGGGGGTGGTGTGGCCCTGGAGGGAGGCCATGACTTG 2461
Db 989 GlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGluGlyGlyHisAspLeu 1008
Qy 2462 ACCGCCATCTGTGATGCTCTGAGCTTGTCTCTGGCTCTGCTCAGTGTAAAGCTGCAG 2521
Db 1009 ThrAlaIleCysAspAlaSerGluAlaCysValSerAlaLeuLeuSerValGluLeuGln 1028
Qy 2522 CCCTTGGATGAGGAGCTCTGCAGCAAAAGCCACACATCAACGAGCGAGCGCCACGCTAG 2581
Db 1029 ProteaspGluAlaValLeuGlnGlnLysProAsnIleAsnAlaValAlaThrLeuGlu 1048
Qy 2582 AAAGTCATCGAGATCCAGAGCAAAACACTGGAGCTGTGTGCAGAAAGTTCGCCGCTGGTCTG 2641
Db 1049 LysValIleGluIleGlnSerLysHisTrpSerCysValGlnLysPheAlaAlaGlyLeu 1068
Qy 2642 GGCCGGTCCCTCGCAGGGGGCCCAAGCAGGTGAGACCGAAGCGCAAT -GTGAAGCC 2700
Db 1069 GlyArgSerLeuArgGluAlaGlnAlaGlyGluThrGluGluAlaGluThrValSerAla 1088
Qy 2701 ATGGCTTGTCTGTGGTGGGGCCGACAGCCCAAGCTGGCGAGCGCCGCGGACACAGC 2760
Db 1089 MetalLeuLeuSerValGlyAlaGluGlnAlaGlnAlaAlaAlaAlaArgGluHisSer 1108
Qy 2761 CCCAGGGCGGAGGAGCCCATCGAGCAGGAGCGCTGCCCTG 2802
Db 1109 ProArgProAlaGluGluProMetGluGlnGluProAlaLeu 1122

RESULT 2

US-10-072-094-91
; Sequence 91, Application US/10072094
; Publication No. US2003002538A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, DONALD
; APPLICANT: LORENZI, MATTHEW
; APPLICANT: ATTAR, RICARDO
; APPLICANT: GOTTARDIS, MARCO
; TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
; FILE REFERENCE: 3053-4145US1
; CURRENT APPLICATION NUMBER: US/10/072,094
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,296
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 91
; LENGTH: 1122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-094-91

Alignment Scores: 7.84e-228 Length: 1122
Pred. No.: 4793.00 Matches: 927
Score:

Percent Similarity: 99.57% Conservative: 3
Best Local Similarity: 89.25% Mismatches: 4
Query Match: 89.66% Indels: 1
DB: 9 Gaps: 0
US-09-502-945-2 (1-2885) x US-10-072-094-91 (1-1122)
Qy 2 GAATTCCTCTGTGCAAGTCAAGAGGACCCACACAGCGCGCCTCAACCATTCCTCCCA 61
Db 189 GluPheLeuLeuSerLysSerLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 208
Qy 62 CAGCACCCCAAAATCCTGGGAGGCCACCATGCTCTTTTGGACACAGAGTTCCTCCCTCCCA 121
Db 209 GlnHisProLysCysTrpGlyAlaHisHisAlaSerLeuAspGlnSerSerProGln 228
Qy 122 AGCGGCCCCCTGGAGCGCTCCCTCTACAACTGCCTTTGGCTGGCGCCTACGACAGT 181
Db 229 SerGlyProGlyThrProProSerTyrLysLeuProLeuProGlyProTyrAspSer 248
Qy 182 CGAGACACTTCCCTCCCGCAAAACAGCCTCTGAACCCAACTTGAAGTGCCTTCAAG 241
Db 249 ArgAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg 268
Qy 242 CTAATAACAGAGGTGGTGGAGCGAGAGAGTCCCTCTGCTGCTGCAAGGATGGACT 301
Db 269 LeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArgLysAspGlyThr 288
Qy 302 GTTATTAGCACCTTTAAGAAGAGAGAGTGTGTGAGATCATCAGGTCCCGGGCTGGGCGT 361
Db 289 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyAlaGlyProGlyAlaSer 308
Qy 362 TCCTGTGTGTAAACAGCGACCCGCTCCGCGCCAGCTCTCCCAACAGCTCCCAACAGCACC 421
Db 309 SerValCysAsnSerAlaProGlySerGlyProSerSerProAsnSerSerHisSerThr 328
Qy 422 ATCGCTGAGAAATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCAG 481
Db 329 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGln 348
Qy 482 CACCGACCTCTCTCTGGACAGCTCCCAACAGCTTACAGCTCTACAGCTCTCTTCT 541
Db 349 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTyrThrSerProSer 368
Qy 542 CTGCCCCAACATCTCCCTAGGGCTGACGGCCACGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
Db 369 LeuProAsnIleSerLeuGluAlaThrValThrValThrValThrValThrValThrVal 388
Qy 602 GCCTCCCGCAAGCTGTGACACAGAGGAGCGGAGGAGCGGAGGAGGAGGAGGAGGAGGAGG 661
Db 389 AlaSerProLysLeuSerThrGlnGlnGluAlaGluArgGlnAlaLeuGlnSerLeuArg 408
Qy 662 CAGGGTGGCAGCTGACCGCAAGTTCATGAGCACATCTCTATTCTGCTGCTGCTGCTGCTG 721
Db 409 GlnGlyGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCysLeuLeu 428
Qy 722 GCGTGGCAGCTGAGGGGACGAGCGGCCCGCCAGGAGCTCCCTGCTGCTGCTGCTGCTGCTG 781
Db 429 GlyValAlaLeuGluGlyAspGlySerProHisGlyHisAlaSerLeuLeuGlnHisVal 448
Qy 782 CTGTTGCTGGAGAGCGGCCGACAGACACCTCATTTGCTGCTGCTGCTGCTGCTGCTGCTG 841
Db 449 LeuLeuLeuGluGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 468
Qy 842 TCCCCACTAGTAGCGGTGAACGTGTGGCCACAGCATGCGGACGCTAGCAAGCTCCCG 901
Db 469 SerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLysLeuPro 488
Qy 902 CGGCATCGGCCCTGAGCGCAGCTCAGTCTCACCGCTGCCGAGAGTCCCCAGGCGCTG 961
Db 489 ArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGlnAlaLeu 508
Qy 962 CAGCAGCTGGTCAACAACACAGCAGCTCTCTGAGAGAGCAGAGCAGCAGCAGCAGCAG 1021
Db 1021 CAGCAGCTGGTCAACAACACAGCAGCTCTCTGAGAGAGCAGAGCAGCAGCAGCAGCAG 1021

Db	509	GlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnLysGlnGlnGln	528
Qy	1022	CTACAGTGGCGAAGATCTCTACCAAGACAGAGGAGCTGCCACGACGACCCACCCAC	1081
Db	529	LeuGlnLeuGlyLysIleLeuThrLysThrGlyGluLeuProArgGlnProThrThrHis	548
Qy	1082	CCTGAGGAGACAGAGAGAGCTGACGAGACAGACAGAGAGTCTTGTCTGGGGGAGGAGCC	1141
Db	549	ProGluGluThrGluGluLeuThrGluGlnGlnGluValLeuLeuGlyGluGlyAla	568
Qy	1142	CTGACCATGCCCGGAGGGCTCCACAGAGAGTGAGACACACAGGAAGACCTGGAGGAG	1201
Db	569	LeuThrMetProArgGluGlySerThrGluSerGluSerThrGlnGluAspLeuGluGlu	598
Qy	1202	GAGGACGAGGAAGAGATGGGAGGAGGAGGAGGATTGCATCCAGGTAAAGCAGGAGGAG	1261
Db	589	GluAspGluGluAspGlyGluGluGluAspCysIleGlnValLysAspGluGlu	608
Qy	1262	GCGGAGAGTGGTCTCAGAGAGGGGCCGACTTGGAGGAGCCTGGTGTGATACAAAAA	1321
Db	609	GlyGluSerGlyAlaGluGluGlyProAspLeuGluGluProGlyAlaGlyTyLysLys	628
Qy	1322	CTGTCTTCAGATCCCAACCGCTGCAACTTTCAGAGTGTTACCAAGCGCCCTCAGCGCTG	1381
Db	629	LeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyGlnAlaProLeuSerLeu	648
Qy	1382	GCCACTGTGCCCCACCAACCGCTGGGCCCTACCCAATCCTCCCTGTGCTCCCTGGGGC	1441
Db	649	AlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaProGlyGly	668
Qy	1442	ATGAAGAACCCCCAGACCAACCGTCAAGCACCTCTTCCACCAAGTGTGGTCTAGCAC	1501
Db	669	MetLysSerProProAspGlnProValLysHisLeuPheThrThrGlyValValTyAsp	688
Qy	1502	ACGTTTCATGCTAAAGCACACAGTCATGTGCGGGACACACACAGTCGACCCTGACGTGCT	1561
Db	689	ThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGluHisAla	708
Qy	1562	GGCGGATCCAGAGCATCTGGTCCCGGCTGCAGGAGACAGCGCTGCTAGCAAGTGGCAG	1621
Db	709	GlyArgIleGlnSerIleIrpSerArgLeuGlnGluThrGlyLeuLeuSerLysCysGlu	728
Qy	1622	CGGATCCGAGGTTCGAAAGCCACGCTAGATGATGATCCAGACAGTGCATCTGGAATACCAC	1681
Db	729	ArgIleArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGluTyHis	748
Qy	1682	ACCTGCTCTATCGGACCAGTCCCTCAACCGGACAGAGCTAGACAGCAAGATTGCTC	1741
Db	749	ThrLeuLeuTyGlyThrSerProLeuAsnArgGlnLysLeuAspSerLysLysLeuLeu	768
Qy	1742	GGTCCCATCCAGCAAGATGTATGCTGTGCTGCTTGTGGGGGCATCGGGTGGACAGT	1801
Db	769	GlyProIleSerGlnLysMetTyAlaValLeuProCysGlyGlyIleGlyValAspSer	788
Qy	1802	GACACCTGTGTGAATGAGATGCACTCCTCCAGTGTGTGCGCATGGCAGTGGGTGCTGCTG	1861
Db	789	AspThrValTrpAsnGluMetHisSerSerSerAlaValArgMetAlaValGlyCysLeu	808
Qy	1862	CTGAGCTGGCCCTTCAGGTGGCTGCAGAGAGCTCAAGAATGGATTGGCCATCATCCGG	1921
Db	809	LeuGluLeuAlaPheLysValAlaAlaGlyLeuLysAsnGlyPheAlaIleArg	828
Qy	1922	CCCCGAGCACCCAGCCGAGGAATCCACAGCATGGGATTCGTCTCTCAACTCTGTA	1981
Db	829	ProProGlyHisHisAlaGluGluSerThrAlaMetGlyPheCysPhePheAsnSerVal	848
Qy	1982	GCCATCACCCGAAACTCTCTACAGCAGAAGTTGAACGTGGCAGAGGCTCCTCATCGTGGAC	2041
Db	849	AlaIleThrAlaLysLeuLeuGlnLysLeuAsnValGlyLysValLeuIleValAsp	868
Qy	2042	TGGGACATTCACCATGCGAATGCCACCCAGAGCGGTCTTACAATGACCCCTCTGTGTC	2101
Db	869	TrpAspIleHisHisGlyAsnGlyThrGlnGlnAlaPheTyTrpAsnAspProSerValLeu	888

Qy	2102	TACATCTCTCTGCATCGCTATGACAAACGGGAACTTCTTTTCAGGCTCTGGGGCTCCTGAA	21161
Db	889	TyrIleSerLeuHisArgTyrAspAsnGlyAsnPhePheProGlySerGlyAlaProGlu	908
Qy	2162	GAGGTTGGTGAGACAGCGCTGGGGTACAATGTGAACGTGGCATGGACAGAGAGTGTG	22211
Db	909	GluValGlyGlyGlyProGlyValGlyTyrAsnValAsnValAlaThrThrGlyGlyVal	928
Qy	2222	GACCCCCCAATTGGAGACGTGGAGTACCTTACAGCCTTTCAGGACAGCTGGTGTATGCCCCAAT	22811
Db	929	AspProIleGlyAspValGluTyrLeuThrAlaPheArgThrValValMetProIle	948
Qy	2282	GGCCACAGAGTTCACCTGATGTGTGTCTAGTCTCCGCGGGTGTGATGCTGTGTGAAGGA	23411
Db	949	AlaHisGluPheSerProAspValLeuValSerAlaGlyPheAspAlaValGluGly	968
Qy	2342	CATCTGTCCTCTCGGTGGCTACTCTGTGCACCGCCAGATGTTTGGCCACTTGCACCAAG	24011
Db	969	HisLeuSerProLeuGlyGlyTyrSerValThrAlaArgCysPheGlyHisLeuThrArg	988
Qy	2402	CAGCTGATGACCTTGGCAGGGCGGGTGTGTCTGGCCCTGGAGGGAGGCCATGACTTG	24611
Db	989	GlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGluGlyGlyHisAspLeu	1008
Qy	2462	ACGCCCATGTGTGATGCCCTGAGCTTGTGTCTCGGCTCTGCTCAGTGTAAAGCTCCAG	25211
Db	1009	ThrAlaIleCysAspAlaSerGluAlaCysValSerAlaLeuLeuSerValGluLeuGln	1028
Qy	2522	CCCTTGATCAGCAGCTCTTTCAGCAAAAGCCACATCAACGAGTGGCCACGCTAGAG	25811
Db	1029	ProLeuAspGluAlaValLeuGlnGlnLysProAsnIleAsnAlaValAlaThrLeuGlu	1048
Qy	2582	AAAGTCATCAGATCCAGAGCAAACTGGAGCTGTGTGCAGAGTTCCGCGCTGGTCTG	26411
Db	1049	LysValIleGluIleGlnSerLysHisIlePheSerCysValGlnLysPheAlaAlaGlyLeu	1068
Qy	2642	GGCCGGTCCCTGCGAGGGGCCCAAGCAGGTGAGACCGAAGACCGCAAAAT-GTGAAGCCC	27001
Db	1069	GlyArgSerLeuArgGluAlaGlnAlaGlyGluThrGluGluAlaGluThrValSerAla	1088
Qy	2701	ATGCCCTTGTCTGTGTGGGGCCGAACAGGCCCAAGCTGGCGGACGCCGGGAACACAGC	27601
Db	1089	MetAlaLeuLeuSerValGlyAlaGluGlnAlaGlnAlaAlaAlaArgGluHisSer	1108
Qy	2761	CCAGGCGCGCAGAGGAGCCATGGAGCAGGAGCCTGGCCCTG	2802
Db	1109	ProArgProAlaGluGluProMetGluGlnGluProAlaLeu	1122
RESULT 3			
US-09-817-913-9			
; Sequence 9, Application US/09817913			
; Patent No. US20020061860A1			
; GENERAL INFORMATION:			
; APPLICANT: Li, Zuomei			
; APPLICANT: Bonfils, Claire			
; APPLICANT: Besterman, Jeffrey			
; TITLE OF INVENTION: Inhibition of Specific Histone Deacetylase Isoforms			
; FILE REFERENCE: 106101.145			
; CURRENT APPLICATION NUMBER: US/09/817,913			
; CURRENT FILING DATE: 2001-03-26			
; PRIOR APPLICATION NUMBER: US 60/192,157			
; PRIOR FILING DATE: 2000-03-24			
; NUMBER OF SEQ ID NOS: 33			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 9			
; LENGTH: 717			
; TYPE: PRT			
; ORGANISM: Human			
US-09-817-913-9			
Alignment Scores:			
Pred. No.:			
5.5e-170			
Length: 717			

Score: 3612.50 Matches: 706
 Percent Similarity: 98.75% Conservative: 3
 Best Local Similarity: 98.33% Mismatches: 6
 Query Match: 67.57% Indels: 4
 DB: 10 Gaps: 3

US-09-502-945-2 (1-2885) x US-09-817-913-9 (1-717)

QY 656 CTGGGAGGCTGGACGCTGACCGGCAAGTTCATGAGCACATCCTCTATTCTGGCTGC 715
 Db 1 LeuArgGlnGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCys 20
 QY 716 CTCGTGGGCTGGCACTGGAGGCGAGCGGAGCCGCCACGGCATGCTCCCTGCTGCAG 775
 Db 21 LeuLeuGlyValAlaLeuGluGlyAspGlySerProHisGlyHisAlaSerLeuLeuGln 40
 QY 776 CATGTGCTGTGTGGAGAGGCGCGGACGACGAGACCCCTCATTTGCTGGCCACTCCAC 835
 Db 41 HisValLeuLeuLeuGluGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHis 60
 QY 836 GGCAGTCCCACTAGTACGGGTGAACGTGTGGCCACCAAGCATGGCGAGGTAGGCAAG 895
 Db 61 GlyGlnSerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLys 80
 QY 896 CTCGCGGGCATCGGCCCTGAGCCGCACTAGTCTCCTACCGTGGCGCAGACTCCCCAG 955
 Db 81 LeuProArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGln 100
 QY 956 GCCTGTGAGCAGCTGGTCAATCAACACACACAGCAGTCTCCTACCGTGGCGCAGAGCAG 1015
 Db 101 AlaLeuGlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnLysGln 120
 QY 1016 CAGCAGCTACAGCTGGGCAAGATCCTCACCAAGACAGGGAGCTGCCAGGCGCCACCC 1075
 Db 121 GlnGlnLeuGlnLeuGlyLysIleLeuThrLysThrGlyGluLeuProArgGlnProThr 140
 QY 1076 ACCACCTGAGGACAGAGGAGGAGCTGACGAGCAGCAGGAGTCTTGTGGGGGAG 1135
 Db 141 ThrHisProGluGluThrGluGluGluLeuThrGluGlnGlnValLeuLeuGlyGlu 160
 QY 1136 GGAGCCTGACCATGCCCGGAGGCTCCACAGAGAGTGAGACACACAGGAAGACCTG 1195
 Db 161 GlyAlaLeuThrMetProArgGluGlySerThrGluSerGluSerThrGlnGluAspLeu 180
 QY 1196 GAGGAGGAGGACGAGGAGGAGTGGGAGGAGGAGGATTCATCCAGGTAAAGGAC 1255
 Db 181 GluGluGluAspGluGluAspGly---GluGluGluAspCysIleGlnValLysAsp 199
 QY 1256 GAGGAGGCGAGAGTGTGCTGAGGAGGGCCCGACTTGGAGGAGCTGGTGGTGGATAC 1315
 Db 200 GluGluGlySerGlyAlaGluGluGlyProAspLeuGluGluProGlyAlaGlyTyr 219
 QY 1316 AAAAACTGTTCTCAGATGCCCAACCGCTGCAACCTTTCAGGTGTACCAAGCGCCCTC 1375
 Db 220 LysLysLeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyrGlnAlaProLeu 239
 QY 1376 AGCCTGGCCACTGTGCCCAACCAAGCCCTGGGCGGTACCAATCCCTCCCTGCTGCCCT 1435
 Db 240 SerLeuAlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaPro 259
 QY 1436 GGGGGATGAGAACCCCCCAGACCAACCGCTCAAGCACCTCTTCAACCAAGTGTGGTC 1495
 Db 260 GlyGlyMetLysSerProProAspGlnProValLysHisLeuPheThrThrGlyValVal 279
 QY 1496 TAGCACACCTTATGCTTAAGACACCATGTCATGTGGGGAGAACACACACACCTGAG 1555
 Db 280 TyrAspThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGlu 299
 QY 1556 CATGCTGGCGGATCCAGAGCATCTGGTCCCGCTGCAGACAGGCGCTGTAGCAAG 1615
 Db 300 HisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnGluThrGlyLeuLeuSerLys 319
 QY 1616 TCGAGCGGATCCGAGTCCGAAAGCCACGCTAGATGATCCACACAGTGCACCTCTGNA 1675

Db 320 CysGluArgIleArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGlu 339
 QY 1676 TAC---CACACCCTGCTCTATGGACCAGTCCCTCAACCGCAGAAAGTAGACAGCAAG 1732
 Db 340 TyrIleHisThrLeuLeuTyrGlyThrSerProLeuAsnArgGlnLysLeuAspSerLys 359
 QY 1733 AAGTTGCTCGGTCCCATCAGCCAGCAAGATGTATGTGTGCTGCTCTGTGGGGCATCGGG 1792
 Db 360 LysLeuLeuGlyProIleSerGlnLysMetTyrAlaValLeuProCysGlyIleGly 379
 QY 1793 GTGGCAGTGACACCGCTGTGGAATGAGATGCACCTCCTCCAGTGTGTGGCGCATGCAGTG 1852
 Db 380 ValAspSerAspThrValTrpAsnGluMetHisSerSerSerAlaValAlaArgMetAlaVal 399
 QY 1853 GCGTGCCTCTGGAGCTGGCCTTCAAGTGTGCTGAGGAGAGCTCAAGAATGGATTGCTC 1912
 Db 400 GlyCysLeuLeuGluLeuAlaPheLysValAlaAlaGlyGluLeuLysAsnGlyPheAla 419
 QY 1913 ATCATCCGCGCCCGGACACCGCAGGAGGATCCACAGCCATGGGATTCTGCTCTTC 1972
 Db 420 IleIleArgProProGlyHisHisAlaGluGluSerThrAlaMetGlyPheCysPhe 439
 QY 1973 AACTCTGTAGCCATCACCGCAAACTCCTACAGCAGAAAGTTGAACGTGGCAAGTCTCTC 2032
 Db 440 AsnSerValAlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeu 459
 QY 2033 ATCTGGAGCTGGGACATTCACCATGGCAATGGCACCCAGCAGCGCTTCTACAATGACCCC 2092
 Db 460 IleValAspTrpAspIleHisHisGlyAsnGlyThrGlnGlnAlaPheTyrAsnAspPro 479
 QY 2093 TCTGTGCTCTACATCTCTGTCATCGCTATGACACGGGAACTTCTTCCAGGCTCTGGG 2152
 Db 480 SerValLeuTyrIleSerLeuHisArgTyrAspAsnGlyAsnPheProGlySerGly 499
 QY 2153 GCTCTGTGAAGAGTGTGGTGGAGGACCGGTGGGTACAATGTCAACGTGGCATGGACA 2212
 Db 500 AlaProGluGluValGlyGlyGlyProGlyValGlyTyrAsnValAsnValAlaTrpThr 519
 QY 2213 GAGGTGTGGACCCCGCATTTGGAGACGTGGAGTACCTTACAGCCTTCAGACAGTGTG 2272
 Db 520 GlyGlyValAspProIleGlyAspValGluTyrLeuThrAlaPheArgThrValVal 539
 QY 2273 ATGCCATTGCCACGAGTCTCCTCCTGATGTGTC---CTAGTCTCCGCGGTTTGTAT 2329
 Db 540 MetProIleAlaHisGluPheSerProAspValValThrLeuValSerAlaGlyPheAsp 559
 QY 2330 GCTGTGAAGGACATCTCTCTCTCTGGTGGCTACTCTGTCCACCGCCAGATGTTTGGC 2389
 Db 560 AlaValGluGlyHisLeuSerProLeuGlyTyrSerValThrAlaArgCysPheGly 579
 QY 2390 CACTTGACCGACGCTGATACCTGGCAGGGGCGGGTGGTGTGCTGGCCCTGGAGGGA 2449
 Db 580 HisLeuThrArgGlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGluGly 599
 QY 2450 GCCCATGACTTGACCGCCATCTGTGATGCCCTCTGAAGCTTGTCTCGGCTCTGCTCAGT 2509
 Db 600 GlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValSerAlaLeuLeuSer 619
 QY 2510 GTAAAGCTGCAGCCCTTGGATGAGGACGCTTTCAGCAAAAAGCCCAACATCAACGCACTG 2569
 Db 620 ValGluLeuGlnProLeuAspGluLeuValLeuGlnGlnLysProAsnIleAsnAlaVal 639
 QY 2570 GCCAGCTTAGAAGATCATCGATCCAGAGCAAACTGGAGCTGTGTGTGCAAGATTC 2629
 Db 640 AlaThrLeuGluLysValIleGluThrGlnSerLysHisTrpSerCysValGlnLysPhe 659
 QY 2630 GCCGCTGTGTGGCGGCTCCCTCGGAGGGGCCCAAGCAGGTGAGACCGAAGAGCCGAA 2689
 Db 660 AlaAlaGlyLeuGlyArgSerLeuArgGluAlaGlnAlaGlyGluThrGluGluAlaGlu 679
 QY 2690 AT-GTGAACGCCATGGCCCTTCTGTGTGGTGGGCGCCGACAGGCCAAGCTGCGCGACCC 2748
 Db 2748

Db 680 ThrValSerAlaMetAlaLeuLeuSerValGlyAlaGluGlnAlaGlnAlaAlaAla 699
Qy 2749 CGGAGACACCCAGCCGCGCAGAGAGCCCATGGACGAGGACCTGCCCTG 2802
|||||
Db 700 ArgGluHisSerProArgProAlaGluProMetGluGlnGluProAlaLeu 717
RESULT 4
US-09-817-538-9
; Sequence 9, Application US/09817538
; Patent No. US20020137162A1
; GENERAL INFORMATION:
; APPLICANT: Li, Zuomei
; APPLICANT: Bonfils, Claire
; APPLICANT: Besterman, Jeffrey
; TITLE OF INVENTION: Antisense Oligonucleotide Inhibition of Specific Histone
; TITLE OF INVENTION: Deacetylase Isoforms
; FILE REFERENCE: 106101.144
; CURRENT APPLICATION NUMBER: US/09/817,538
; CURRENT FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/192,157
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 9
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Human
US-09-817-538-9
Alignment Scores:
Pred. No.: 5,5e-170 Length: 717
Score: 3612.50 Matches: 706
Percent Similarity: 98.75% Conservative: 3
Best Local Similarity: 98.33% Mismatches: 6
Query Match: 67.57% Indels: 4
DB: 10 Gaps: 3
US-09-502-945-2 (1-2885) x US-09-817-538-9 (1-717)
Qy 656 CTGCGCGAGGTGGCAGCTGACCGCAGTTCATGACACATCTCTATTCTGCTGCG 715
Db 1 LeuArgGlnGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCys 20
Qy 716 CTGCTGGGCGTGGCACTGAGGGCGACGGAGCCCGCCAGCGCATGCTCCCTGCTGCGAG 775
Db 21 LeuLeuGlyValAlaLeuGluGlyAspGlySerProHisGlyHisAlaSerLeuLeuGln 40
Qy 776 CATGTGCTGTGCTGGAGCAGCCCGCGCAGCAGACACCTCATGCTGTGCCACTCCAC 835
Db 41 HisValLeuLeuGluGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHis 60
Qy 836 GGGCAGTCCCCACTAGTGACGGGTGAAGTGTGGCCACGAGCATCGGACGCTAGGCAAG 895
Db 61 GlyGlnSerProLeuValThrGlyGluArgValAlaThrSerMetArgThrValGlyLys 80
Qy 896 CTCCCGCGCATCGGCCCTGAGCGCACTCAGTCTCACCCGCTGCGCAGAGTCCCCAG 955
Db 81 LeuProArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGln 100
Qy 956 GCCTCGCAGAGTGTGTATGCAACACAGCAGCAGCAGCAGTCTCTGGAGAGCAGAGCAG 1015
Db 101 AlaLeuGlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGluLysGlnLysGln 120
Qy 1016 CAGCAGTACAGTGGGCAAGATCTCTACCAACAGCAGGAGGAGCTGCCAGGAGCCAC 1075
Db 121 GlnGlnLeuGlnLeuGlyLysIleLeuThrLysThrGlyGluLeuProArgGlnProThr 140
Qy 1076 ACCCACTCTGAGGAGACAGAGGAGGAGTGTGACGAGCAGCAGAGTCTTGTGGGGAG 1135
Db 141 ThrHisProGluGluThrGluGluLeuThrGluGlnGlnGluValLeuLeuGlyGlu 160
Qy 1136 GGAGCCCTGACCATGCCCGGGAGGGCTCCACAGAGAGTGTGAGAGCAGACAGGAGACCT 1195
|||||

Db 161 GlyAlaLeuThrMetProArgGluGlySerThrGluSerGluSerThrGlnGluAspLeu 180
Qy 1196 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1255
|||||
Db 181 GluGluGluAspGluGluAspGly---GluGluGluAspCysIleGlnValLysAsp 199
Qy 1256 GAGGAGGCGAGAGTGTGCTGAGGAGGCGCGACTTGGAGGAGGCTGGTGTGGATAC 1315
Db 200 GluGluGlyGluSerGlyAlaGluGluGlyProAspLeuGluGluProGlyAlaGlyTyr 219
Qy 1316 AAAAACTGTCTCAGATGCCCAACCGCTGCAACCTTTGAGGTGTACCAAGGCGCCCTC 1375
Db 220 LysLysLeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyrGlnAlaProLeu 239
Qy 1376 AGCTGCGCCACTGTGCCCAACCGCTGCGCGCTGACCAATCTCCCTGCTGCCCT 1435
Db 240 SerLeuAlaThrValProHisGlnAlaLeuGlyArgThrGlnSerSerProAlaAlaPro 259
Qy 1436 GGGGCGATGAAGAACCCCCCAGACCAACCGCTCAAGCACCTCTTACCACAAAGTGTGGTC 1495
Db 260 GlyGlyMetLysSerProProAspGlnProValLysHisLeuPheThrThrGlyValVal 279
Qy 1496 TAGACACGTTTATGCTAAAGCACCAGTGCATGTGCGGGGAACACACACGTCACCTGAG 1555
Db 280 TyrAspThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGlu 299
Qy 1556 CATGCTGGCGGATCCAGAGCATCTGTGCCGCTGCGGAGCAGGAGCAGCCTTACCAAG 1615
Db 300 HisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnGluThrGlyLeuLeuSerLys 319
Qy 1616 TGGGAGCGGATCCGAGTGCACAAAGCCACGCTAGATGAGATCCAGACAGTGCCTCTGAA 1675
Db 320 CysGluArgIleArgGlyArgLysAlaThrLeuAspGluIleGlnThrValHisSerGlu 339
Qy 1676 TAC---CACACCTGCTCTATGGACCCAGTCCCTCAACCGCAGAGAGTAGCAGCAAG 1732
Db 340 TyrIleHisThrLeuLeuTyrGlyThrSerProLeuAsnArgGlnLysLeuAspSerLys 359
Qy 1733 AAGTGTGCTGCTCCATCAGCCAGAGATGTATGCTGTGCTGCTGTGGGGCATCGGG 1792
Db 360 LysLeuLeuGlyProIleSerGlnLysMetTyrAlaValLeuProCysGlyGlyIleGly 379
Qy 1793 GTGACAGTAGTACACCGTGTGGAATGAGATGAGATGCTCTCCAGTGTGCTGCGCAGTGCAGT 1852
Db 380 ValAspSerAspThrValTrpAsnGluMetHisSerSerAlaValArgMetAlaVal 399
Qy 1853 GGCTGCTGCTGGAGTGGCTTCAAGTGGCTGCGCAGAGAGCTCAAGATGGATTGGCC 1912
Db 400 GlyCysLeuLeuGluLeuAlaPheLysValAlaAlaGlyGluLeuLysAsnGlyPheAla 419
Qy 1913 ATCATCGGCGCCCGACGACACCGCGGAGGAAATCCACAGCATGGGATTTGCTTCTTC 1972
Db 420 IleIleArgProProGlyHisHisAlaGluGluSerThrAlaMetGlyPheCysPhePhe 439
Qy 1973 AACTCTGTAGCCATCACCGCAAACTCTCTACAGCAGAGTGAACGTGGGCAAGTGCCTC 2032
Db 440 AsnSerValAlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeu 459
Qy 2033 ATCTGTGACTGGGACATTCACCATGGCAATGGCACCGCAGGCGCTTCTACAATGACCCC 2092
Db 460 IleValAspTrpAspIleHisHisGlyAsnGlyThrGlnGlnAlaPheTyrAsnAspPro 479
Qy 2093 TCTGTGCTCTACATCTCTCTCATCGTATGACAAACGGGAACCTTTTCCAGGCTCTGGG 2152
Db 480 SerValLeuTyrIleSerLeuHisArgTyrAspAsnGlyAsnPhePheProGlySerGly 499
Qy 2153 GCTCCTGAAGAGTGTGGAGGACCGGCTGGGTGTACATGTGTGAACCTGGCATGGACA 2212
Db 500 AlaProGluGluValGlyGlyProGlyValGlyTyrAsnValAsnValAlaTrpThr 519
Qy 2213 GGAGGTGTGGACCCCGCCCATTTGGAGACGTGGAGTACCTTACAGCTTCAGCAGAGTGGTG 2272
Db 520 GlyGlyValAspProIleGlyAspValGluTyrLeuThrAlaPheArgThrValVal 539